



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΠΟΛΙΤΙΚΩΝ ΜΗΧΑΝΙΚΩΝ

**ΤΟΜΕΑΣ ΥΔΑΤΙΚΩΝ ΠΟΡΩΝ ΚΑΙ
ΠΕΡΙΒΑΛΛΟΝΤΟΣ**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

**Πιθανοτικές κατανομές ακραίων
βροχοπτώσεων**

Εφαρμογή σε παγκόσμια κλίμακα

Νεραντζάκη Σοφία

Επιβλέπων : Δ. Κουτσογιάννης, Καθηγητής ΕΜΠ

Αθήνα , Οκτώβριος 2012

Ευχαριστίες

Αρχικά θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή της διπλωματικής μου εργασίας κ. Δημήτρη Κουτσογιάννη για τις γνώσεις που μου παρείχε και το ενδιαφέρον που έδειξε κατά τη διάρκεια της εκπόνησής της. Η εργασία αυτή δε θα μπορούσε να πραγματοποιηθεί χωρίς τη καθοδήγηση και την έμπνευση του Σίμωνα Παπαλεξίου τον οποίο ευχαριστώ θερμά για τη στήριξη και το κουράγιο του.

Ευχαριστώ επίσης την Πετρούλα για τη θετική της ενέργεια και τη βοήθειά της στα τελευταία στάδια της εργασίας μου, καθώς και τον Κώστα για τις συμβουλές του και την ηθική συμπαράσταση. Ακόμα, τις φίλες μου Δανάη, Εύα και Άνυ για το ενδιαφέρον που έδειξαν για την πρόοδό μου καθώς και τη βοήθεια που μου παρείχαν στη διάρκεια των φοιτητικών μου χρόνων.

ΠΕΡΙΕΧΟΜΕΝΑ

ΠΕΡΙΛΗΨΗ	vii
ABSTRACT	ix
1 ΕΙΣΑΓΩΓΗ	1
1.1 ΑΝΤΙΚΕΙΜΕΝΟ ΤΗΣ ΕΡΓΑΣΙΑΣ	1
1.2 ΔΙΑΡΘΡΩΣΗ ΤΗΣ ΕΡΓΑΣΙΑΣ	3
2 ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΕΠΙΣΚΟΠΗΣΗ	4
2.1 ΕΙΣΑΓΩΓΙΚΑ	4
2.2 ΙΣΤΟΡΙΚΗ ΑΝΑΔΡΟΜΗ ΣΤΗ ΜΕΘΟΔΟΛΟΓΙΑ ΑΚΡΑΙΩΝ ΤΙΜΩΝ	5
2.3 ΕΙΣΑΓΩΓΗ ΣΤΗ ΘΕΩΡΙΑ ΑΚΡΑΙΩΝ ΤΙΜΩΝ	7
2.4 ΘΕΩΡΗΜΑ FISHER-TIPPETT	8
2.5 ΓΕΝΙΚΕΥΜΕΝΗ ΚΑΤΑΝΟΜΗ ΑΚΡΑΙΩΝ ΤΙΜΩΝ	10
2.5.1 Πεδία έλξης κατανομών	12
• Τύπος ορίου Fréchet	13
• Τύπος ορίου αντίστροφης Weibull	14
• Τύπος ορίου Gumbel.....	14
2.6 ΓΕΝΙΚΕΥΜΕΝΗ ΚΑΤΑΝΟΜΗ PARETO	15
2.7 ΤΑΞΙΝΟΜΗΣΗ ΚΑΤΑΝΟΜΩΝ ΒΑΣΕΙ ΣΥΜΠΕΡΙΦΟΡΑΣ ΟΥΡΑΣ	17
2.7.1 Γενικά.....	17
2.7.2 Κατά Werner και Upper	17
2.7.3 Κατά Embrechts et al.	21
2.7.4 Κατά Ouarda et al.....	23
3 ΑΝΑΛΥΣΗ	27

3.1 ΕΙΣΑΓΩΓΗ	27
3.1.1 Στατιστική Προσέγγιση.....	27
3.1.2 Ορισμός Ανώτερου Μέρους Κατανομής	27
3.1.3 Σχεδιαστικές Θέσεις.....	28
3.2 ΥΠΟ ΕΞΕΤΑΣΗ ΤΥΠΟΙ ΚΑΤΑΝΟΜΩΝ	29
3.2.1 Εκθετική Κατανομή	29
3.2.2 Κατανομή Pareto τύπου II.....	29
3.2.3 Λογαριθμοκανονική Κατανομή	30
3.2.4 Κατανομή Weibull	32
3.2.5 Κατανομή Γάμμα	33
3.3 ΥΠΟ ΕΞΕΤΑΣΗ ΟΥΡΕΣ ΚΑΤΑΝΟΜΩΝ	34
3.3.1 Ουρά Κατανομής Pareto τύπου II.....	34
3.3.2 Ουρά Λογαριθμοκανονικής Κατανομής	35
3.3.3 Ουρά Κατανομής Weibull.....	36
3.3.4 Ουρά Κατανομής Γάμμα.....	37
4 ΜΕΘΟΔΟΛΟΓΙΑ	39
4.1 ΔΕΔΟΜΕΝΑ.....	39
4.2 ΓΡΑΦΙΚΗ ΠΑΡΑΣΤΑΣΗ ΣΕ ΛΟΓΑΡΙΘΜΙΚΟΥΣ ΑΞΟΝΕΣ (LOG-LOG PLOT)	40
4.2.1 Θεωρητικό Υπόβαθρο	40
4.2.2 Εφαρμογή μεθόδου σε βασικές θεωρητικές κατανομές.....	41
4.2.3 Εφαρμογή Μεθόδου στα δεδομένα	44
4.2.4 Προσομοίωση Monte Carlo	45
4.3 ΜΕΘΟΔΟΣ MEAN EXCESS FUNCTION (Συνάρτηση Μέσης Υπέρβασης).....	48
4.3.1 Θεωρητικό υπόβαθρο.....	48

4.3.2 Εφαρμογή μεθόδου σε βασικές θεωρητικές κατανομές.....	50
4.3.3 Προσομοίωση Monte Carlo	53
4.3.4 Εφαρμογή Μεθόδου στα δεδομένα.....	65
4.4 ΜΕΘΟΔΟΣ ΕΛΑΧΙΣΤΟΠΟΙΗΣΗΣ ΣΦΑΛΜΑΤΟΣ	69
4.4.1 Σύγκριση Τύπων ελαχιστοποίησης σφάλματος	69
4.4.2 Εφαρμογή Μεθόδου στα δεδομένα.....	75
5 ΣΥΝΟΨΗ ΚΑΙ ΣΥΜΠΕΡΑΣΜΑΤΑ.....	79
6 ΒΙΒΛΙΟΓΡΑΦΙΚΕΣ ΑΝΑΦΟΡΕΣ	83
ΠΑΡΑΡΤΗΜΑ Α	87
ΠΑΡΑΡΤΗΜΑ Β	103

ΠΕΡΙΛΗΨΗ

Η ουρά μιας κατανομής, που είναι το ανώτερο μέρος αυτής, καθορίζει το μέγεθος και τη συχνότητα των ακραίων γεγονότων. Οι πιθανοτικές κατανομές, με βάση τη συμπεριφορά της ουράς τους, μπορούν να ταξινομηθούν σε δύο βασικές οικογένειες: τις κατανομές με «βαριές» ουρές και τις κατανομές με «ελαφριές», με τις δεύτερες να έχουν πιο ήπια και με μικρότερη συχνότητα ακραία γεγονότα. Στην περίπτωση της βροχόπτωσης, η μελέτη της ουράς εξάγει σημαντικά συμπεράσματα για την περίοδο επαναφοράς των ακραίων τιμών της και είναι προφανής η συμβολή της στο σωστό υδρολογικό σχεδιασμό.

Προκειμένου να αξιολογηθεί η συμπεριφορά των ακραίων της βροχόπτωσης, εξετάζονται 3 477 σταθμοί σε όλο τον κόσμο, που έχουν μήκος δείγματος άνω των 100 ετών. Σε αυτούς εφαρμόζονται δύο διαδοχόμενες γραφικές μέθοδοι με σκοπό μια πρώτη κατάταξη της ουράς τους στις προαναφερθείσες οικογένειες. Η πρώτη μέθοδος είναι η γραφική παράσταση ουράς σε λογαριθμικούς άξονες, η οποία ελέγχει αν η ουρά είναι τύπου δύναμης (βαριά ουρά). Ωστόσο, η προσομοίωση Monte Carlo απέδειξε ότι η μέθοδος αυτή είναι αναξιόπιστη για ακραίες τιμές. Η δεύτερη μέθοδος είναι η Συνάρτηση Μέσης Υπέρβασης (Mean Excess Function), που εκφράζει τη μέση τιμή πάνω από ένα όριο μιας μεταβλητής και το γράφημά της έχει μηδενική κλίση για την Εκθετική κατανομή (οικογένεια λεπτών ουρών). Για την εφαρμογή της μεθόδου κατασκευάστηκαν διαστήματα εμπιστοσύνης για την κλίση της Εκθετικής κατανομής συναρτήσει του μήκους δείγματος. Η επικύρωση της μεθόδου με την προσομοίωση Monte Carlo απέδειξε ότι αυτή δίνει καλά αποτελέσματα ιδιαίτερα για μεγάλα δείγματα. Επιπρόσθετα, γίνεται σύγκριση μεταξύ τεσσάρων νορμών μέσου τετραγωνικού σφάλματος, με σκοπό την εύρεση της πιο αξιόπιστης για την προσαρμογή κατανομής σε ουρά. Χρησιμοποιώντας την αποτελεσματικότερη προσαρμόζονται τέσσερις θεωρητικές κατανομές ουράς στις εμπειρικές ουρές. Οι θεωρητικές κατανομές είναι: η Pareto τύπου II, που είναι κατανομή δύναμης, η Λογαριθμοκανονική, που έχει επίσης βαριά ουρά, η Weibull, που έχει ενδιάμεση και η Γάμμα που έχει λεπτή ουρά και χρησιμοποιείται συχνά στη μελέτη ημερήσιας βροχόπτωσης. Τα αποτελέσματα από την προσαρμογή συνάδουν σε ικανοποιητικό βαθμό με αυτά από τη μέθοδο της Σ.Μ.Υ. Η ανάλυση αποδεικνύει ότι οι κατανομές με βαριές ουρές συμφωνούν καλύτερα με τα μέγιστα των βροχοπτώσεων σε σχέση με τις συχνά χρησιμοποιούμενες κατανομές με ελαφριές ουρές.

ABSTRACT

The tail of a probability distribution, which is the upper part of it, governs both the magnitude and the frequency of extreme events. The probability distributions can be generally categorized into two families, based on their tail behavior: heavy-tailed and light-tailed distributions, with the latter having milder and less frequent extremes. In the case of rainfall, the study of the tail draws important conclusions concerning the return period of extreme events and its contribution in hydrologic design is obvious.

In order to evaluate the behavior of rainfall extremes, 3 477 stations from all over the world with sample size over 100 years are examined. Two common graphical methods are applied in order to classify the empirical tails of rainfall to the above families. The first method, the log-log plot, demonstrates whether the tail is a power-type (heavy-tailed); however, a Monte Carlo study revealed that the method is unreliable for extreme values. The second method, the Mean Excess Function, is based on the mean value of a variable over a threshold and results in a zero slope line when applied and graphically depicted for the Exponential distribution (light-tailed). To apply the method we constructed confidence intervals for the slope of the Exponential distribution as functions of sample size. The validation of the method using Monte Carlo techniques revealed that it performs well especially for large samples. Additionally, four mean square error norms are compared in order to find the most reliable for fitting the distribution tail. The best performing norm is used to fit four different theoretical distribution tails to the empirical tails. The theoretical distributions tails are: the Pareto type II which is a power type, the Lognormal which is also heavy-tailed, the Weibull which has an intermediate tail, and the Gamma which is light-tailed and commonly used to describe daily rainfall. The results from this approach are well-matched with the ones obtained by the application of the Mean Excess Function method. The analysis shows that heavy-tailed distributions are in general in better agreement with the rainfall extremes compared to the commonly used light-tailed ones.

1 ΕΙΣΑΓΩΓΗ

1.1 ΑΝΤΙΚΕΙΜΕΝΟ ΤΗΣ ΕΡΓΑΣΙΑΣ

Στα πλαίσια της εργασίας αυτής εξετάζονται δεδομένα βροχόπτωσης από 3 477 σταθμούς σε όλο τον κόσμο. Συγκεκριμένα, η βροχόπτωση αντιμετωπίζεται ως τυχαία μεταβλητή που ορίζεται από μία κατανομή και εφαρμόζονται μέθοδοι για την εύρεση κάθε φορά κατάλληλης κατανομής που να προσαρμόζεται ικανοποιητικά στις ακραίες τιμές βροχόπτωσης. Έχοντας δεδομένη την κατανομή που ακολουθούν οι ακραίες τιμές, είμαστε σε θέση να αντιστοιχίσουμε κάθε ακραία τιμή βροχόπτωσης με μια περίοδο επαναφοράς και να συμπεράνουμε αν ένα γεγονός βροχόπτωσης είναι πράγματι ακραίο.

Οι ακραίες τιμές μιας κατανομής αποτελούν το ανώτερο μέρος αυτής, που είναι γνωστό ως ουρά της κατανομής. Η ουρά ορίζει το μέγεθος και τη συχνότητα των ακραίων γεγονότων. Σε γενικές γραμμές, η συμπεριφορά της ουράς όλων των πιθανοτικών κατανομών μπορεί να ταξινομηθεί σε δύο βασικές οικογένειες: τις κατανομές με «βαριές» ή «χονδρές» ουρές και τις κατανομές με «ελαφριές» ή «λεπτές» ουρές, με τις δεύτερες να έχουν πιο ήπια και με μικρότερη συχνότητα ακραία γεγονότα. Εδώ εξετάζονται τα δεδομένα βροχής με σκοπό τόσο τη γενική τους κατηγοριοποίηση στις δύο αυτές οικογένειες, όσο και την ανεύρεση συγκεκριμένης κατανομής που να εφαρμόζεται καλά στις ακραίες τιμές τους.

Η συνήθης διαδικασία για την επιλογή κατανομής βροχόπτωσης στο σύνολο των δεδομένων είναι

- (α) Η εκ των προτέρων επιλογή κάποιων παραμετρικών οικογενειών κατανομών, έχοντας ως αναφορά το μεγάλο ιστορικό χρήσης οικογενειών κατανομών ανάλογα με τον ερευνητικό τομέα.
- (β) Η εκτίμηση των παραμέτρων σύμφωνα με μια από τις μεθόδους προσαρμογής κατανομών
- (γ) Η επιλογή της καταλληλότερης κατανομής σύμφωνα με κάποια τεστ εφαρμογής.

Όμως σ' αυτήν ακριβώς την επιλογή της οικογένειας είναι που μπορεί να γίνει σημαντικό λάθος. Συγκεκριμένα, είναι δύσκολο να πραγματοποιηθεί η επιλογή μιας οικογένειας που θα αποτελεί καλό μοντέλο για τις ουρές της κατανομής. Μόνο μικρό τμήμα από τα εμπειρικά

δεδομένα ανήκει στην ουρά της κατανομής (εκτός και αν διατίθεται πολύ μεγάλο δείγμα) και όλοι οι μέθοδοι προσαρμογής κατανομών που αφορούν στο σύνολο των δεδομένων θα μεροληπτούν εις βάρος της ουράς, εφόσον οι εκτιμώμενες παράμετροι θα υπολογίζονται κυρίως για το μεγαλύτερο τμήμα των δεδομένων (που εξ ορισμού δεν ανήκει στην ουρά). Προφανώς, μια κακή προσαρμογή στην ουρά έχει ως αποτέλεσμα σημαντικά λάθη και επιπτώσεις στον υδρολογικό σχεδιασμό. Σ' αυτό το στάδιο φαίνεται η χρησιμότητα της Θεωρίας Ακραίων Τιμών.

Η μελέτη λοιπόν της ουράς μετεωρολογικών και υδρολογικών φαινομένων, και συγκεκριμένα της βροχόπτωσης, με τη συνεπαγόμενη ανεύρεση της περιόδου επαναφοράς, αποτελεί σημαντική γνώση για τον υδρολογικό σχεδιασμό και την κατασκευή μεγάλων υδραυλικών έργων, μέσω της εκτίμησης του πλημμυρικού κινδύνου. Ο σχεδιασμός των αντιπλημμυρικών έργων γίνεται για συγκεκριμένη περίοδο επαναφοράς και ο υδρολογικός σχεδιασμός περιλαμβάνει δύο συνιστώσες: την εκτίμηση καταιγίδας σχεδιασμού και την εκτίμηση της πλημμύρας που προκαλείται από αυτή (Ευστρατιάδης 2011). Η βροχόπτωση, σε συνδυασμό με λοιπές υδρολογικές διεργασίες (εξατμισοδιαπνοή, κατακράτηση, διήθηση, επιφανειακή και υπόγεια απορροή) και το γεωγραφικό και γεωλογικό ανάγλυφο, οδηγεί στην τελική διαμόρφωση της παροχής.

Παράλληλα, η παραπάνω διαδικασία παρουσιάζει κλιματολογικό ενδιαφέρον, καθώς με τον τρόπο αυτό επιτυγχάνεται πρόβλεψη της μακροπρόθεσμης ακραίας βροχόπτωσης. Μάλιστα, το γεγονός ότι πραγματοποιείται σε παγκόσμια κλίμακα δίνει τη δυνατότητα σύγκρισης της συμπεριφοράς της βροχόπτωσης σε διάφορα σημεία του κόσμου. Τέλος, λόγω του μεγάλου αριθμού των προς επεξεργασία σταθμών πραγματοποιήθηκε συστηματοποίηση των μεθόδων και οι προτάσεις αυτές προσφέρονται για βελτίωση ήδη υπάρχοντων προγραμμάτων που στόχο έχουν την επιλογή κατάλληλης κατανομής σε δεδομένα.

1.2 ΔΙΑΡΘΡΩΣΗ ΤΗΣ ΕΡΓΑΣΙΑΣ

Η εργασία περιλαμβάνει, εκτός από την παρούσα εισαγωγή (Κεφάλαιο 1), πέντε Κεφάλαια και δύο Παραρτήματα.

Στο Κεφάλαιο 2 πραγματοποιείται βιβλιογραφική επισκόπηση της Θεωρίας Ακραίων Τιμών και συγκεντρώνονται κάποιοι βασικοί τρόποι κατατάξης ουρών κατανομών.

Στο Κεφάλαιο 3 αναφέρονται και αναλύονται οι βασικές κατανομές και οι αντίστοιχες ουρές κατανομών οι οποίες ελέγχθηκαν.

Στο Κεφάλαιο 4 αναπτύσσεται η μεθοδολογία που ακολουθήθηκε τόσο για τη διαμόρφωση εικόνας για το «πάχος» της ουράς των δεδομένων όσο και για την προσαρμογή κατανομών στα μέγιστα αυτών.

Στο Κεφάλαιο 5 συνοψίζονται τα αποτελέσματα των παραπάνω μεθόδων και εξάγονται συμπεράσματα.

Στο Κεφάλαιο 6 παρατίθενται οι βιβλιογραφικές αναφορές

Στο Παράρτημα Α δίνονται βασικά γραφήματα από τους σταθμούς με τα περισσότερα δεδομένα, που διασαφηνίζουν τη συμπεριφορά της ουράς της βροχόπτωσης και αναφέρονται ξεχωριστά σε κάθε μέθοδο.

Στο Παράρτημα Β παρατίθεται ο κώδικας που χρησιμοποιήθηκε σε κάθε μέθοδο (στο πρόγραμμα Wolfram Mathematica 8)

2 ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΕΠΙΣΚΟΠΗΣΗ

2.1 ΕΙΣΑΓΩΓΙΚΑ

Η Θεωρία των ακραίων τιμών (EVT) είναι ένας κλάδος της στατιστικής που έχει ως αντικείμενο την ανάπτυξη και μελέτη στοχαστικών μοντέλων με σκοπό την επίλυση προβλημάτων που σχετίζονται με την εμφάνιση «πολύ μεγάλων» ή «πολύ μικρών» τιμών και άρα ακραίων και σπάνιων γεγονότων σε τυχαία πειράματα, σε αντίθεση με την κλασική στατιστική την οποία ενδιαφέρει η μέση συμπεριφορά της στοχαστικής διαδικασίας. Οι μέθοδοι ακραίων τιμών είναι στατιστικές μέθοδοι που αναλύουν, όπως είναι εμφανές, τις ακραίες τιμές μιας διεργασίας.

Η Θεωρία Ακραίων Τιμών μπορεί να χωριστεί σε δύο μέρη:

- (α) Το πιθανοθεωρητικό μέρος, που αφορά τη μελέτη της στοχαστικής συμπεριφοράς ακραίων παρατηρήσεων
- (β) Το στατιστικό μέρος, που αποτελείται από μοντέλα ερμηνείας, εκτιμήσεων και προβλέψεων ακραίων συμβάντων με βάση πραγματικά δεδομένα.

Η Θεωρία Ακραίων Τιμών βρίσκει εφαρμογές τόσο στον τομέα των οικονομικών, όσο και στην Υδρολογία αναφορικά με την προστασία έναντι πλημμυρών (π.χ. με την πρόβλεψη της μέγιστης ροής που αναμένεται μία φορά στα 100 έτη) καθώς επίσης και στην Μετεωρολογία για την πρόβλεψη ακραίων φαινομένων όπως ακραίοι άνεμοι, για την αξιολόγηση του κινδύνου σε σταθμούς παραγωγής ενέργειας, για γεγονότα ισχυρής βροχόπτωσης, για κύματα καύσωνα, τυφώνες και ξηρασίες και για ακραία φαινόμενα σε ένα μεταβαλλόμενο κλίμα.

Για να βρεθεί η οριακή κατανομή μεγίστων (και ελαχίστων) υπάρχουν δύο τάξεις κατανομών ακραίων τιμών. Η πρώτη ονομάζεται Γενικευμένη Κατανομή Ακραίων Τιμών, εισήχθη από τον Jenkinson (1955) και περιλαμβάνει τρεις κατανομές ακραίων τιμών. Η δεύτερη ονομάζεται Γενικευμένη Κατανομή Pareto, εισήχθη από τους Balkema και de Haan (1974) και τον Pickands (1975), και περιλαμβάνει την κατανομή Pareto, την Εκθετική και την Ομοιόμορφη. Και οι δύο τάξεις αναλύονται παρακάτω.

2.2 ΙΣΤΟΡΙΚΗ ΑΝΑΔΡΟΜΗ ΣΤΗ ΜΕΘΟΔΟΛΟΓΙΑ ΑΚΡΑΙΩΝ ΤΙΜΩΝ

Οι πρώτες έρευνες στον τομέα της Θεωρίας Ακραίων Τιμών έγιναν από στατιστικούς αναλυτές. Έπειτα ασχολήθηκαν οι υδρολόγοι στις αρχές του 1970, οι αναλογιστές το 1980, οι αναλυτές οικονομικών χρονοσειρών στα μέσα του 1990 και τελικά οι διαχειριστές ποσοτικού κινδύνου στα τέλη του 1990. Παρακάτω δίνονται τα κυριότερα ιστορικά γεγονότα στη Θεωρία Ακραίων Τιμών (Dalle Molle, 2002)

Η τυπική κατανομή Pareto πήρε το όνομά της από τον Ιταλό μηχανικό και οικονομολόγο Vilfredo Pareto (1848-1923). Ο Pareto ήταν ο πρώτος που μελέτησε δυναμικού τύπου κατανομές και για το λόγο αυτό οι κατανομές αυτές αναφέρονται συχνά ως κατανομές τύπου Pareto. Ο Pareto ερεύνησε τα στατιστικά χαρακτηριστικά του πλούτου των ατόμων σε μια σταθερή οικονομία.

Οι πρώτες πρωτοποριακές έρευνες πάνω στη Θεωρία Ακραίων Τιμών έγιναν από τον Fisher (1927) και τους Fisher και Tippett (1928) και αφορούσαν την ασυμπτωτική συμπεριφορά των κατανομών μεγίστου και ελαχίστου του δείγματος. Ήταν οι πρώτοι που διατύπωσαν τους τρεις θεμελιώδεις τύπους ασυμπτωτικών κατανομών ακραίων τιμών. Το Θεώρημα Fisher-Tippett περιγράφει την ασυμπτωτική συμπεριφορά του δείγματος της κατανομής των μεγίστων και είναι ανάλογο με το Κεντρικό Οριακό Θεώρημα για το δείγμα της κατανομής των κανονικοποιημένων αθροισμάτων (Fisher & Tippett, 1928).

Η κατανομή Weibull πήρε το όνομά της από τον Σουηδό φυσικό Waloddi Weibull ο οποίος πρώτος πρότεινε αυτή την κατανομή το 1939 στο πλαίσιο των ερευνών για την αντοχή υλικών (Weibull, 1939)

Ο von Mises (1936) πρώτος πρότεινε το συνδυασμό των τριών τύπων κατανομών της Θεωρίας Ακραίων Τιμών σε έναν τύπο που ονομάζεται Γενικευμένη Κατανομή Ακραίων Τιμών. Ο ίδιος έδωσε τις επαρκείς συνθήκες για να ανήκει μια κατανομή στο μέγιστο πεδίο έλξης μιας κατανομής ακραίων τιμών (1936) οι οποίες συχνά καλούνται συνθήκες von Mises (von Mises, 1936).

Ο Gnedenko (1943) ήταν ο πρώτος που εξέτασε τους τρεις θεμελιώδεις οριακούς τύπους ακραίων τιμών από αυστηρά μαθηματική σκοπιά. Από αυτόν δόθηκαν οι αναγκαίες και ικανές συνθήκες για να ανήκει μια ακολουθία τυχαίων μεταβλητών σε κάποιο από τα τρία

πεδία έλξης (1943) και έπειτα ο de Haan έδωσε σε αυτές μια πιο απλή μορφή (1976). Ο Gnedenko επίσης έδειξε ότι αν η ουρά της κατανομής φθίνει σαν συνάρτηση δύναμης, τότε η κατανομή ανήκει στο πεδίο έλξης της κατανομής Fréchet.

Η εφαρμοσιμότητα της Θεωρίας Ακραίων Τιμών μελετήθηκε εκτενώς από τον Gumbel (Gumbel, 1958). Πρότεινε την ομώνυμη κατανομή καθώς και τη μέθοδο Μεγίστων Τεμαχίων (Block Maxima), η οποία είναι ένα από τα παλαιότερα μοντέλα της κλασσικής ανάλυσης ακραίων τιμών (ονομάζεται αλλιώς Μέθοδος Υπερβάσεων Gumbel). Οι κατανομές των μεγίστων και ελαχίστων συνδέονται με την αρχή της ιδιότητας της συμμετρίας, η οποία εισήχθη από τον Gumbel. Η αρχή αυτή υποδηλώνει ότι τα αποτελέσματα των μεγίστων μπορούν να τροποποιηθούν κατάλληλα για να εφαρμοστούν στα ελάχιστα. Ο Gumbel τέλος παρουσιάζει τις λεπτομέρειες διαδικασιών στατιστικής εκτίμησης και δίνει πολλά επεξηγηματικά παραδείγματα εφαρμογών Θεωρίας Ακραίων Τιμών στη Φυσική και τη Μηχανική.

Η γ – παραμετροποίηση των κατανομών ακραίων τιμών οφείλεται στον von Mises (1936) και συχνά συνδυάζεται και με το όνομα του Jenkinson (1955).

Η Γενικευμένη Κατανομή Ακραίων Τιμών με τρεις παραμέτρους εισήχθη από τον Jenkinson (1955 και 1969) για να προσδιορίσει τις κατανομές συχνότητας των μεγαλύτερων τιμών μετεωρολογικών δεδομένων όταν είναι άγνωστη η οριακή μορφή της κατανομής ακραίων τιμών.

Το θεώρημα Pickands-Balkema-de Haan, το οποίο συνδέει την οριακή μορφή της κατανομής υπέρβασης για υψηλά κατώφλια με τη συνάρτηση γενικευμένης κατανομής Pareto, απεδείχθη από τους Balkema και de Haan (1974) και από τον Pickands (1975). Απέδειξαν ότι η οριακή έκφραση για τις διαβαθμισμένες υπερβάσεις της τυχαίας μεταβλητής X πάνω από ένα υψηλό κατώφλι u μπορεί να προσεγγιστεί από τη Γενικευμένη Pareto κατανομή.

Η μέθοδος Άνω Κατωφλίου (Peaks over Threshold - POT) χρησιμοποιείται από τους υδρολόγους για πάνω από 25 χρόνια (Embrechts et al., 1997)

Οι πρώτες εφαρμογές στη διαχείριση κινδύνου έγιναν με τη συμβολή των αναλογιστών Andrew J. McNeil και Paul Embrechts και των συνεργατών τους στο Ελβετικό Πανεπιστήμιο ETH στη Ζυρίχη (1998 έως 2005)

2.3 ΕΙΣΑΓΩΓΗ ΣΤΗ ΘΕΩΡΙΑ ΑΚΡΑΙΩΝ ΤΙΜΩΝ

Σύμφωνα με τη Θεωρία Ακραιών Τιμών είναι γνωστό ότι για δεδομένο αριθμό n ανεξάρτητων, ταυτοτικά κατανεμημένων τυχαίων μεταβλητών, η μεγαλύτερη από αυτές $X = \max \{Y_1, Y_2, \dots, Y_n\}$ ακολουθεί συνάρτηση πιθανότητας

$$H_n(x) = [F(x)]^n \quad (2.1)$$

$$\text{Αφού } H_n(x) = P(X \leq x) = P[\max(Y_1, Y_2, \dots, Y_n) \leq x] = P(Y_1 \leq x, \dots, Y_n \leq x) = [F(x)]^n$$

Όπου $F(x) = P(Y_i \leq x)$ είναι η συνάρτηση πυκνότητας πιθανότητας των Y_i και αναφέρεται συνήθως ως μητρική κατανομή. Εάν το n δεν είναι σταθερό, αλλά μπορεί να θεωρηθεί ως πραγματοποίηση μιας τυχαίας μεταβλητής με κατανομή Poisson και μέση τιμή ν , τότε η κατανομή της μεταβλητής X γίνεται (Todorovic & Zelenhasic 1970; Rossi et al., 1984) σύμφωνα με Koutsoyiannis 2004

$$H'_\nu(x) = \exp\{-\nu[1-F(x)]\} \quad (2.2)$$

$$\text{Όμως: } \ln[F(x)]^n = n \ln\{1-[1-F(x)]\} = n\{-[1-F(x)] - [1-F(x)]^2 - \dots\} \approx -n[1-F(x)]$$

κάνοντας χρήση του αναπτύγματος Taylor όπου $\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \dots$ και θέτοντας

όπου x το $-[1-F(x)]$. Άρα έχουμε ότι για μεγάλα n ή μεγάλα $F(x)$, $H_n(x) \approx H'_n(x)$.

Αριθμητική έρευνα δείχνει ότι ακόμα και για σχετικά μικρά n , η διαφορά μεταξύ $H_n(x)$ και $H'_n(x)$ δεν είναι σημαντική.

Στις υδρολογικές εφαρμογές που αφορούν την κατανομή των ετήσιων μεγίστων βροχής ή πλημμύρας, μπορεί να υποθεθεί ότι ο αριθμός των τιμών του Y_i (π.χ. ο αριθμός των μεγίστων κατακρημνίσεων ή των παροχών πλημμύρας) δεν είναι σταθερός, άρα το μοντέλο Poisson μπορεί να θεωρηθεί κατάλληλο γι' αυτή την περίπτωση. Δεδομένης και της μικρής διαφοράς μεταξύ των $H_n(x)$ και $H'_n(x)$, μπορούμε να συμπεράνουμε ότι η εξίσωση (2.2) είναι κατάλληλο μοντέλο για κάθε πρακτική υδρολογική εφαρμογή.

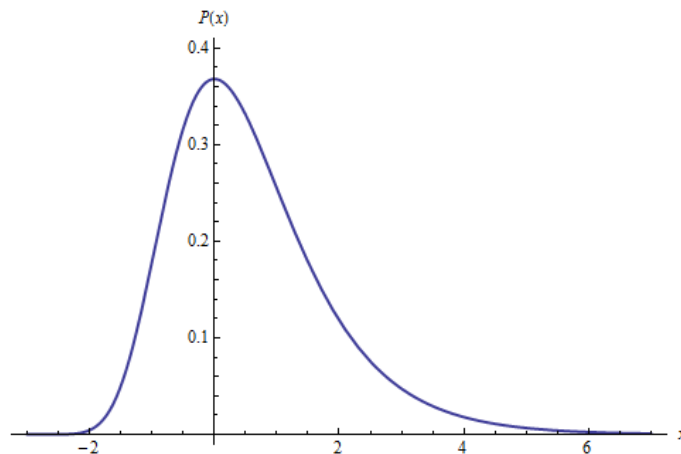
Όμως, οι εξισώσεις (2.1) και (2.2), των οποίων η αξιολόγηση απαιτεί να είναι γνωστή η μητρική κατανομή, δε χρησιμοποιούνται στην υδρολογική στατιστική. Αντί γι αυτές, στον τομέα αυτό έχουν χρησιμοποιηθεί ασυμπτωτικές ή οριακές κατανομές ακραίων τιμών, οι οποίες λαμβάνονται από τις παραπάνω εξισώσεις όταν το n τείνει στο άπειρο. Ουσιαστικά, δε βασιζόμαστε στην F , αλλά προσεγγίζουμε την κατανομή για μεγάλο n , ανεξαρτήτως της F .

2.4 ΘΕΩΡΗΜΑ FISHER-TIPPETT

Σύμφωνα με το θεώρημα Fisher-Tippett (1928), αν το μέγιστο του ορθά κανονικοποιημένου δείγματος συγκλίνει σε μια μη εκφυλισμένη κατανομή, τότε ανήκει σε μια από τις παρακάτω τρεις κατανομές (αθροιστική συνάρτηση κατανομής)

(α) Τύπος 1, Κατανομή τύπου Gumbel (Σχήμα 2.1)

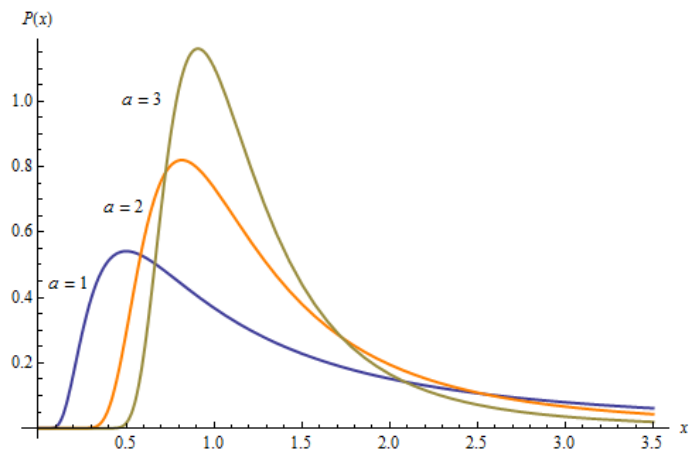
$$\Lambda(x) = P(X \leq x) = e^{-e^{-x}}, \quad x \in \mathbf{R} \quad (2.3)$$



Σχήμα 2.1 Διάγραμμα συνάρτησης πυκνότητας πιθανότητας κατανομής τύπου Gumbel

(β) Τύπος 2, Κατανομή τύπου Fréchet (Σχήμα 2.2)

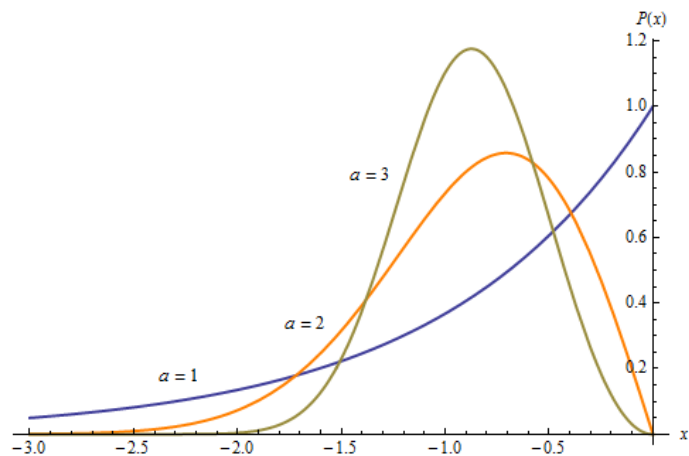
$$\Phi_a(x) = P(X \leq x) = \begin{cases} 0 & x \leq 0 \\ e^{-x^{-a}} & x > 0 \end{cases}, \quad a > 0 \quad (2.4)$$



Σχήμα 2.2 Διάγραμμα συνάρτησης πυκνότητας πιθανότητας κατανομής τύπου Fréchet για διάφορες τιμές του δείκτη α .

(γ) Τύπος 3, Αντίστροφη Weibull Κατανομή (Σχήμα 2.3)

$$\Psi_a(x) = P(X \leq x) = \begin{cases} e^{-(-x)^a} & x \leq 0 \\ 1 & x > 0 \end{cases}, \quad a > 0 \quad (2.5)$$



Σχήμα 2.3 Διάγραμμα συνάρτησης πυκνότητας πιθανότητας κατανομής τύπου Weibull για διάφορες τιμές του δείκτη α .

Οι κατανομές Φ_a, Ψ_a, Λ καλούνται τυπικές κατανομές ακροτάτων, ενώ κατανομές που είναι του ίδιου τύπου με αυτές καλούνται κατανομές ακροτάτων.

Κατά τον ορισμό, μια τ.μ. Y (ή, αντίστοιχα, η κατανομή της F) ανήκει «στο πεδίο έλξης μιας κατανομής ακροτάτων H » αν υπάρχουν ακολουθίες $c_n > 0, d_n \in \mathbb{R}$ έτσι ώστε $(X - d_n)/c_n \rightarrow_d H$ και συμβολίζουμε Y ή $F \in \text{MDA}(H)$. Με άλλα λόγια, εάν η κατανομή των μεγίστων του δείγματος μιας δεδομένης κατανομής συγκλίνει σε έναν από τους τρεις τύπους, τότε ανήκει στο μέγιστο πεδίο έλξης των $\Lambda(x), \Phi_a(x)$ ή $\Psi_a(x)$.

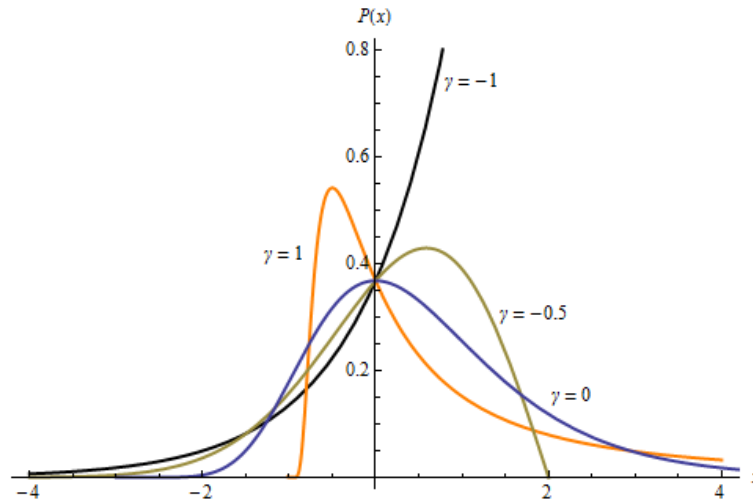
Από μαθηματική άποψη οι τρεις αυτοί τύποι κατανομών συνδέονται. Έστω $X > 0$, τότε αν X έχει σ.κ. $\Phi_a \Leftrightarrow \ln X^a$ έχει σ.κ. $\Lambda \Leftrightarrow -X^{-1}$ έχει σ.κ. Ψ_a (Haan & Ferreira 2006).

2.5 ΓΕΝΙΚΕΥΜΕΝΗ ΚΑΤΑΝΟΜΗ ΑΚΡΑΙΩΝ ΤΙΜΩΝ

Οι von Mises (1936) και Jenkinson (1955), ανέπτυξαν μια συμπυκνωμένη θεωρία κατανομών ακραίων τιμών σύμφωνα με την οποία, και οι τρεις ασύμπτωτες μπορούν να περιγραφούν από μία οικογένεια κατανομών που καλείται κατανομή ΓΑΤ – Γενικευμένη κατανομή Ακραίων Τιμών (GEV – Generalized Extreme Value distribution) και περιγράφεται από τη Σχέση 2.6:

$$H(x) = \exp \left\{ - \left[1 + \gamma \left(\frac{x - \alpha}{\beta} \right)^{-1/\gamma} \right] \right\}, \quad x > \alpha - \beta/\gamma, \quad \gamma > 0 \quad (2.6)$$

όπου $\alpha, \beta > 0$ και $\gamma > 0$ είναι παράμετροι θέσης, κλίμακας και σχήματος αντίστοιχα. Το διάγραμμα συνάρτησης πυκνότητας πιθανότητας της κατανομής φαίνεται στο Σχήμα 2.4.



Σχήμα 2.4 Διάγραμμα συνάρτησης πυκνότητας πιθανότητας κατανομής Γ.Α.Τ. για διάφορες τιμές του δείκτη γ .

Ο Leadbetter (1974) έδειξε ότι αυτό ισχύει όχι μόνο για τα μέγιστα ανεξάρτητων τυχαίων μεταβλητών, αλλά και για εξαρτημένες τυχαίες μεταβλητές, με την προϋπόθεση ότι δεν υπάρχει μακρόχρονη εξάρτηση των υπερβάσεων υψηλών επιπέδων.

Όταν το $\gamma > 0$, η $H(x)$ αντιπροσωπεύει την κατανομή των μεγίστων τύπου II. Σ' αυτή την περίπτωση, η μεταβλητή είναι κάτω φραγμένη και μη άνω φραγμένη. Ειδική περίπτωση έχουμε όταν το κατώτατο όριο γίνει μηδέν ($\alpha = 1/\gamma$). Αυτή η ειδική δι-παραμετρική κατανομή έχει την απλοποιημένη μορφή:

$$H(x) = \exp\left\{-\left(\frac{x-\alpha}{\beta}\right)^{-1/\gamma}\right\}, \quad x \geq \alpha \quad (2.7)$$

Η τελευταία εξίσωση είναι η κατανομή Fréchet (βλ. Σχέση 2.4).

Η οριακή περίπτωση $\gamma = 0$, αντιπροσωπεύει την κατανομή μεγίστου τύπου I ή κατανομή Gumbel (βλ. Σχέση 2.3). Έτσι, η εξίσωση GEV παίρνει τη μορφή

$$H(x) = \exp\left[-\exp\left(-\frac{x-\alpha}{\beta}\right)\right], \quad x \in \mathbb{R} \quad (2.8)$$

που είναι μη φραγμένη άνω και κάτω ($-\infty < x < +\infty$).

Τέλος, όταν $\gamma < 0$, η $H(x)$ αντιπροσωπεύει τον τύπο III κατανομής μεγίστων ή αντίστροφη Weibull (βλ. Σχέση 2.5) που έχει τη μορφή:

$$H(x) = \exp \left[- \left(- \frac{x - \alpha}{\beta} \right)^{1/\gamma} \right], \quad x < \alpha \quad (2.9)$$

Όμως, ο τύπος αυτός δεν έχει πρακτικό ενδιαφέρον στην υδρολογία, καθώς αναφέρεται σε τυχαίες μεταβλητές άνω φραγμένες όπως αναφέρει. Αν και γενικά η κατηγορία αυτή είναι η συχνότερα χρησιμοποιούμενη στη φύση, η βροχή δεν εμπίπτει σ' αυτή την κατηγορία, καθώς σε σημαντικό βαθμό, οι «ποσότητες» βροχής είναι ανεξέλεγκτες και μπορούν να παρατηρηθούν υψηλές βροχοπτώσεις (Jenkinson, 1955). Επίσης, να παρατηρήσουμε ότι για την κατανομή ελαχίστου, οι ρόλοι των τύπων II και III είναι αντίστροφοι, για παράδειγμα, η κατανομή τύπου III δεν είναι άνω φραγμένη, και γι αυτό είναι λογική επιλογή για τη μελέτη ξηρασιών (από Koutsoyiannis, 2004).

Με τη γ - παραμετροποίηση που έγινε από τους von Mises (1936) και Jenkinson (1955) μια ισοδύναμη έκφραση είναι η Σχέση 2.10. Μέσω αυτής, οι οριακές κατανομές απεικονίζονται ως έχοντες μία παράμετρο, τη γ , η οποία ονομάζεται δείκτης ακραίων τιμών (de Haan & Ferreira, 2000).

$$H(x) = \exp \left[- (1 + \gamma x)^{-1/\gamma} \right], \quad 1 + \gamma x > 0 \quad (2.10)$$

Για διάφορες τιμές της παραμέτρου γ (ή ισοδύναμα της $1/\alpha$) προκύπτουν οι σχέσεις 2.3 έως 2.5.

2.5.1 Πεδία έλξης κατανομών

- *Ορισμοί*

Λέμε ότι μια θετική συνάρτηση h στο $(0, \infty)$ καλείται ομαλής κύμανσης (regularly varying) στο ∞ με δείκτη α και συμβολίζουμε $h \in \mathfrak{R}_\alpha$, αν για κάθε $x > 0$

$$\lim_{x \rightarrow \infty} \frac{h(tx)}{h(x)} = t^\alpha, \quad t \geq 0 \quad (2.11)$$

Η ομαλή κύμανση έχει την έννοια ότι η κατανομή προσομοιάζει με συνάρτηση δύναμης κοντά στο άπειρο, δηλαδή η ουρά της κατανομής είναι βαριά.

Αν $a = 0$ τότε καλείται βραδείας κύμανσης (slowly varying)

Αν $a = -\infty \Leftrightarrow t^a = \begin{cases} \infty, & \text{αν } t \in (0,1) \\ 0, & \text{αν } t \in (1,\infty) \end{cases}$ καλείται ταχείας κύμανσης (rapidly varying)

Κάθε συνάρτηση $h \in \mathfrak{R}_a$, $a \in \mathbb{R}$, μπορεί να γραφεί ως

$$h(x) = x^a L(x) \quad (2.12)$$

όπου L συνάρτηση βραδείας κύμανσης ή $L \in \mathfrak{R}_0$ (Embrechts et al., 1997).

Δύο σ.κ. F, G μπορούν να έχουν διαφορετική μορφή στο \mathbb{R} αλλά η δεξιά ουρά τους να έχει την ίδια συμπεριφορά. Θα καλούμε δύο τέτοιες κατανομές «ισοδύναμης ουράς» (tail-equivalent). Δύο σ.κ. F, G έχουν ισοδύναμη ουρά αν $x_F = x_G$ και, για κάποιο $c > 0$,

$$\lim_{x \rightarrow x_F} \frac{\bar{F}(x)}{\bar{G}(x)} = c \quad (2.13)$$

Αν δύο σ.κ. έχουν ισοδύναμη ουρά, τότε ανήκουν στην ίδια περιοχή έλξης μιας κατανομής ακροτάτων (με τις ίδιες ακολουθίες c_n, d_n).

- **Τύπος ορίου Fréchet**

Οι κατανομές που οδηγούν στο όριο τύπου Fréchet για $\gamma > 0$ χαρακτηρίζονται ως ομαλά κυμαινόμενες. Γενικά ισχύει ότι μια συνάρτηση κατανομής F ανήκει στο μέγιστο πεδίο έλξης (MDA) της Γ.Α.Τ. με $\gamma > 0$ αν και μόνο αν η συνάρτηση ουράς \bar{F} είναι ομαλά κυμαινόμενη. Και γράφουμε (κατά Gnedenko)

$$F \in MDA(\Phi_\alpha) \Leftrightarrow \bar{F}(x) = x^{-1/\gamma} L(x) \quad (2.14)$$

για κάποια συνάρτηση L αργά κυμαινόμενη στο άπειρο (που είναι ισοδύναμη με τη Σχέση 2.12). Αυτό σημαίνει ότι οι κατανομές που δημιουργούν τον τύπο Fréchet είναι κατανομές με

ουρές που είναι συναρτήσεις ομαλά κυμαινόμενες με αρνητικό δείκτη διακύμανσης. Ουσιαστικά οι ουρές τους φθίνουν σε συνάρτηση δύναμης και ο ρυθμός $a = 1/\gamma$ με τον οποίο φθίνουν είναι ο γνωστός δείκτης ουράς της κατανομής.

Κατανομές που συγκλίνουν στη Fréchet είναι Cauchy, Student t, Loggamma, F, Burr και stable (σταθερή, για $a < 2$). Οι κατανομές αυτές παρουσιάζουν το μεγαλύτερο ενδιαφέρον στη Θεωρία Ακραίων Τιμών καθώς έχουν χονδρές ουρές και έχουν υψηλές ροπές που απειρίζονται (McNeil et al., 2005).

- **Τύπος ορίου αντίστροφης Weibull**

Σε αντίθεση με τις κατανομές με βαριές ουρές, η περίπτωση του $\gamma < 0$ περιέχει κατανομές που έχουν πεπερασμένο δεξί ακραίο σημείο $z_+ < \infty$, και για το λόγο αυτό δεν παρουσιάζει ενδιαφέρον για τις υδρολογικές μεταβλητές. Περιλαμβάνει την Ομοιόμορφη και τη Βήτα κατανομή. Ισχύει ότι $F \in MDA(\Psi_a) \Leftrightarrow x_F < \infty$ και $\bar{F}(x_F - x^{-1}) = x^{1/\gamma} L(x)$ για κάποια συνάρτηση L αργά κυμαινόμενη στο άπειρο (McNeil et al., 2005).

- **Τύπος ορίου Gumbel**

Η περίπτωση του $\gamma = 0$ είναι πιο περίπλοκη. Αν και υπάρχουν κι εδώ αναγκαίες και επαρκείς συνθήκες, σπάνια χρησιμοποιούνται στην πράξη. στην κατηγορία αυτή υπάγονται κατανομές όπως η Εκθετική και, αν πούμε ότι η τάξη Fréchet περιέχει οριακούς νόμους για τα μέγιστα του δείγματος κατανομών των οποίων οι ουρές είναι ουσιαστικά τύπου δύναμης, μπορούμε να πούμε ότι η τάξη Gumbel περιέχει οριακούς νόμους για τα μέγιστα του δείγματος κατανομών των οποίων οι ουρές φθίνουν εκθετικά (McNeil et al., 2005).

Μπορεί να αποδειχθεί ότι σε αυτό το πεδίο έλξης υπάρχουν κατανομές με βαριές ουρές, των οποίων όλες οι ροπές είναι πεπερασμένες (όπως για παράδειγμα η Λογαριθμοκανονική κατανομή), αλλά και κατανομές ελαφρών ουρών (όπως η Κανονική, η Εκθετική ή η Γάμμα κατανομή, ή ακόμα κάποιες κατανομές οι οποίες είναι φραγμένες στα δεξιά). Άλλες κατανομές σε αυτή την κατηγορία είναι η X^2 , η Τυπική Weibull (διαφορετική από τη Weibull της GEV), οι Benktander τύπου I και II και η ίδια η Gumbel.

2.6 ΓΕΝΙΚΕΥΜΕΝΗ ΚΑΤΑΝΟΜΗ PARETO

Η προσομοίωση μόνο των μεγίστων μπορεί να είναι ανεπαρκής στην περίπτωση που διατίθενται δεδομένα με μεγάλες χρονοσειρές. Εφόσον η Θεωρία Ακραίων Τιμών στη βάση της ασχολείται με την προσομοίωση της ουράς μιας άγνωστης κατανομής, μια ιδέα είναι η μοντελοποίηση όλων των X_i , των οποίων οι τιμές είναι μεγαλύτερες από ένα σημαντικά υψηλό κατώφλι. Βάσει των αποτελεσμάτων των Balkema και de Haan (1974) και του Pickands (1975), είναι γνωστό ότι αν η κατανομή των Y_i βρίσκεται στο μέγιστο πεδίο έλξης της Γενικευμένης Κατανομής Ακραίων Τιμών, τότε η κατανομή των υπερβάσεων έχει παρόμοια οριακή εκπροσώπηση. Η στενή σχέση μεταξύ της κατανομής μεγίστου $H(x)$ και της ουράς της μητρικής κατανομής $F(x)$ επιτρέπει τον προσδιορισμό της δεύτερης, εάν είναι γνωστή η πρώτη. Η ουρά της $F(x)$ μπορεί να εκπροσωπηθεί από την κατανομή της X με τη δέσμευση να είναι μεγαλύτερη από ορισμένο όριο u , π.χ. $G_u(x) = F(x | x > u)$, για την οποία:

$$1 - G_u(x) = \frac{1 - F(x)}{1 - F(u)}, \quad x \geq u \quad (2.15)$$

Εάν επιλεγεί u τέτοιο ώστε η πιθανότητα υπέρβασης $1 - F(u)$ να ισούται με $1/\nu$, δηλαδή τον αντίστροφο του μέσου αριθμού των γεγονότων σε ένα χρόνο και αν $G(x)$ η υπό συνθήκη κατανομή γι' αυτή τη συγκεκριμένη τιμή, τότε:

$$1 - G(x) = \nu [1 - F(x)] \quad (2.16)$$

Συνδυάζοντας τις εξισώσεις (2.15) και (2.16) παίρνουμε

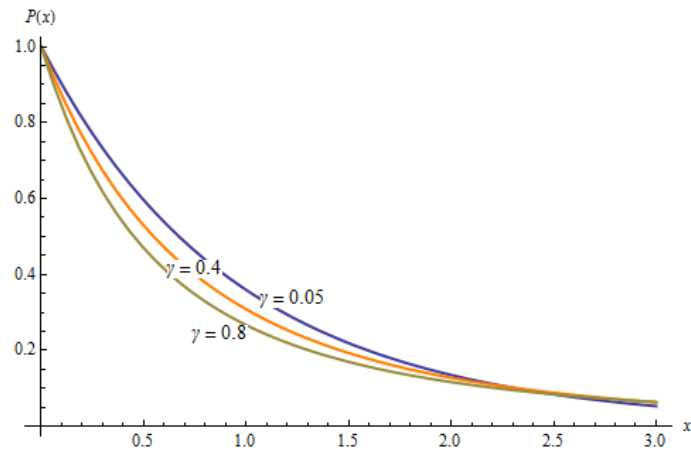
$$G(x) = 1 + \ln H'_\nu(x) \quad (2.17)$$

Εάν δίνεται η $H'_\nu(x)$ από την οριακή κατανομή $H(x)$ στην εξίσωση (2.2), τότε

(α) Για $\gamma > 0$ συνεπάγεται:

$$G(x) = 1 - \left[1 + \gamma \left(\frac{x - \alpha}{\beta} \right) \right]^{-1/\gamma} \quad (2.18)$$

όπου $x \geq a$ όταν $\gamma \geq 0$ και $a \leq x \leq a - \beta/\gamma$ όταν $\gamma < 0$, που είναι η Γενικευμένη κατανομή Pareto (Σχήμα 2.5).

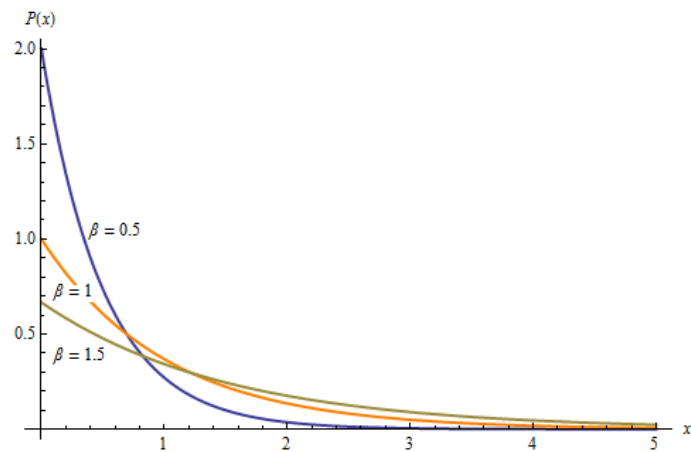


Σχήμα 2.5 Διάγραμμα συνάρτησης πυκνότητας πιθανότητας Γενικευμένης Pareto κατανομής για διάφορες τιμές της παραμέτρου γ ($\alpha = 0$ και $\beta = 1$).

(β) Ομοίως, για $\gamma = 0$:

$$G(x) = 1 - \exp\left(-\frac{x - \alpha}{\beta}\right) \quad (2.19)$$

που είναι η εκθετική κατανομή (Σχήμα 2.6).



Σχήμα 2.6 Διάγραμμα συνάρτησης πυκνότητας πιθανότητας Εκθετικής κατανομής για διάφορες τιμές της παραμέτρου γ .

(γ) Στην περίπτωση που $\gamma = 1$ η κατανομή G είναι η Ομοιόμορφη Κατανομή

$$U[0, \beta] \quad (2.20)$$

Η Γενικευμένη Κατανομή Pareto λοιπόν, προκύπτει ως οριακή κατανομή για υπερβάσεις ενός κατωφλίου αν και μόνο αν η μητρική κατανομή βρίσκεται στο πεδίο έλξης της GEV.

2.7 ΤΑΞΙΝΟΜΗΣΗ ΚΑΤΑΝΟΜΩΝ ΒΑΣΕΙ ΣΥΜΠΕΡΙΦΟΡΑΣ ΟΥΡΑΣ

2.7.1 Γενικά

Συνήθως για να ορίσουμε αν μια ουρά είναι «χονδρή» βασιζόμαστε στην τέταρτη κεντρική ροπή. Έστω X τυχαία μεταβλητή και μ_X και σ_X η μέση τιμή και η τυπική απόκλιση αντίστοιχα, τότε η κατανομή του X λέμε ότι έχει χοντρή ουρά όταν

$$C_k = E \left[\left(\frac{X - \mu_X}{\sigma_X} \right)^4 \right] > 3 \quad (2.21)$$

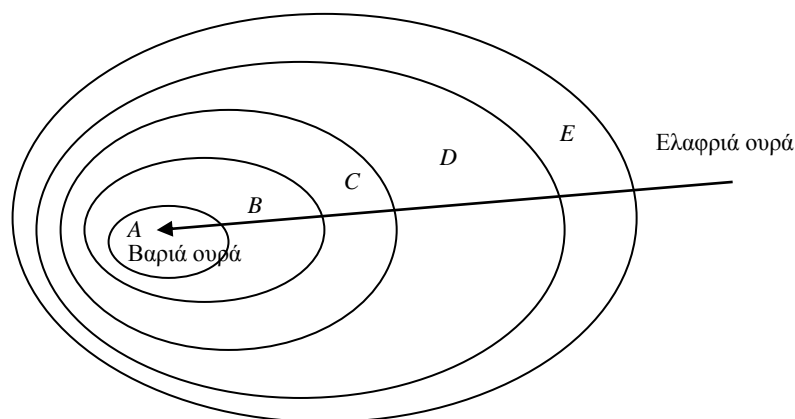
Η $(C_k - 3)$ είναι γνωστή ως επιπλέον κύρτωση λόγω του ότι η τέταρτη κεντρική ροπή (κύρτωση) της Κανονικής Κατανομής είναι ίση με 3. Η κατάταξη όμως με τη Σχέση 2.21 είναι ασαφής και δεν μπορεί να χρησιμοποιηθεί για κατανομές με άπειρη τέταρτη ροπή. Όταν δύο μεταβλητές έχουν άπειρη τέταρτη ροπή, τότε δεν μπορεί να γίνει καμία διάκριση μεταξύ των κατανομών τους με βάση την κύρτωση.

Οι κατανομές μπορούν να κατηγοριοποιηθούν σύμφωνα με την ασυμπτωτική συμπεριφορά της ουράς τους σε δύο γενικές τάξεις, την τάξη των υποεκθετικών, με πυκνότητες πιθανότητας που τείνουν στο μηδέν πιο αργά από την πυκνότητα πιθανότητας της Εκθετικής, και την τάξη των υπερεκθετικών, με πυκνότητες πιθανότητας που τείνουν στο μηδέν ταχύτερα από της Εκθετικής (Teugels 1975; Klüppelberg 1988). Παρακάτω γίνεται αναφορά σε κάποιους από τους τρόπους κατάταξης ουρών που υπάρχουν στη βιβλιογραφία.

2.7.2 Κατά Werner και Upper

Γενικά δεν υπάρχει κάποιος γενικά αποδεκτός ορισμός για τη «βαρύτητα» των ουρών, με τον οποίο θα ήταν δυνατή μια κατάταξή τους. Ένας τρόπος κατάταξης ουρών έχει δοθεί από Werner and Upper (2004), που προτείνουν πέντε τάξεις κατανομών. Οι τάξεις αυτές

εμπεριέχουν η μία την άλλη ($A \subset B \subset C \subset D \subset E$) και μπορούν να παρουσιαστούν όπως φαίνεται στο Σχήμα 2.7 (El Adlouni et al., 2008):



Σχήμα 2.7 Ενδεικτικό σχήμα κατάταξης «χονδρών» ουρών κατά Werner & Upper (2002). Προέλευση: El Adlouni et al. (2008) μετά από προσαρμογή.

Όπου έχουμε τις εξής τάξεις κατανομών:

E : κατανομές με μη ύπαρξη εκθετικών ροπών

D : υπο – εκθετικές κατανομές

C : κατανομές ομαλής κύμανσης

B : κατανομές τύπου Pareto

A : κατανομές σταθεράς – α

Ιδιαίτερης σημασίας είναι οι τάξεις C και B που συνδέονται με την κλασσική θεωρία των ακραίων τιμών.

- E : Η ευρύτερη τάξη E περιλαμβάνει όλες τις κατανομές με $E(e^x) = \infty$. Να σημειωθεί ότι η Κανονική Κατανομή δεν περιλαμβάνεται σ' αυτήν την τάξη καθώς η πιθανότητα της ουράς της $P(X > x) = \bar{F}(x) = 1 - F(x)$ μειώνεται ταχύτερα από εκθετικά. Έτσι, οι κατανομές αυτής της τάξης έχουν χονδρή ουρά σε σχέση με αυτή της Κανονικής.

- Η τάξη D περιλαμβάνει τις υποεκθετικές κατανομές, των οποίων οι ουρές μειώνονται πιο αργά από εκείνες των εκθετικών, και ορίζονται από την ακόλουθη εξίσωση (Teugels, 1975; Goldie and Klüppelberg 1996):

$$\lim_{x \rightarrow \infty} \frac{P(X_1 + \dots + X_n > x)}{P(\max(X_1, \dots, X_n) > x)} = 1 \quad (2.22)$$

Αυτό σημαίνει ότι το άθροισμα n ανεξάρτητων και ισόνομων υποεκθετικών κατανομών είναι πιθανό να υπερβαίνει το x αν και μόνο αν το μέγιστό του είναι μεγαλύτερο από το x . Η εξίσωση (2.22) αποδεικνύεται ότι είναι ισοδύναμη με την (Embrechts et al., 1997) :

$$\lim_{x \rightarrow \infty} \frac{\bar{F}(x)}{\exp(-\varepsilon x)} = \infty, \quad \forall \varepsilon > 0 \quad (2.23)$$

όπου $\bar{F}(x) = 1 - F(x)$ αντιστοιχεί στη συνάρτηση πιθανότητας υπέρβασης και F είναι η αθροιστική συνάρτηση κατανομής. Εφόσον ο παρονομαστής $\exp(-\varepsilon x)$ ($\varepsilon > 0$) αντιπροσωπεύει την ουρά της εκθετικής κατανομής, η τάξη D περιέχει όλες τις κατανομές με ουρές που μειώνονται πιο αργά από τις εκθετικές κατανομές.

- Η τάξη C περιέχει κατανομές οι οποίες αποτελούν υποκατηγορία των υποεκθετικών και ικανοποιούν τη συνθήκη:

$$\lim_{t \rightarrow \infty} \frac{\bar{F}(tx)}{\bar{F}(t)} = x^{-\alpha} \quad (2.24)$$

Η σχέση είναι ισοδύναμη με τη Σχέση 2.11. Η συνθήκη αυτή δηλώνει ότι ασυμπτωτικά, δηλαδή μακριά μέσα στην ουρά, οι πιθανότητες ουράς $\bar{F}(x) = P(X > x)$ μειώνονται σύμφωνα με μια συνάρτηση δύναμης. Συνεπώς, εκεί οι κατανομές συμπεριφέρονται σαν αυτές τύπου Pareto. Οι κατανομές της τάξης C είναι οι ομαλής κύμανσης κατανομές και η παράμετρος α καλείται «δείκτης ουράς» και χρησιμοποιείται ως μέτρο πάχους ουράς.

- Οι κατανομές της τάξης B έχουν ακριβώς ουρές τύπου Pareto. Η αθροιστική συνάρτηση κατανομής της συνάρτησης Pareto είναι:

$$F(x) = 1 - u^\alpha x^{-\alpha} = 1 - \left(\frac{u}{x}\right)^\alpha, \quad x \geq u \text{ και } u > 0 \quad (2.25)$$

όπου $\alpha > 0$ είναι ο δείκτης ουράς. Η συνάρτηση πιθανότητας της ουράς, που είναι με άλλα λόγια η συνάρτηση υπέρβασης πιθανότητας, $\bar{F}(x) = P(X > x)$ της κατανομής της τάξης B είναι επομένως ακριβώς ίση με $u^\alpha x^{-\alpha}$ και όχι ασυμπτωτικά ίση όπως στην περίπτωση της τάξης C . Ο δείκτης ουράς α μπορεί να συσχετιστεί με τις ροπές μιας κατανομής με ουρές Pareto, αφού η συνάρτηση πυκνότητας πιθανότητας μιας τυχαίας μεταβλητής που ακολουθεί κατανομή Pareto είναι $f_{\text{Pareto}}(x) = \alpha u^\alpha x^{-\alpha-1}$ και η ροπή τάξης k δίδεται από:

$$E[X^k] = \alpha u^\alpha \int_u^\infty x^{k-\alpha-1} dx \quad (2.26)$$

Έτσι, μόνο ροπές τάξης k τέτοιες ώστε $k < \alpha$ είναι πεπερασμένες. Αυτή η ιδιότητα φαίνεται χρήσιμη για τον ορισμό της τάξης A . Παρατηρούμε επίσης ότι οι μικρές τιμές του α δίνουν πιο χονδρές ουρές.

- Η τάξη A περιέχει σταθερές κατανομές. Οι σταθερές κατανομές έχουν ουρά Pareto με $\alpha < 2$, που σημαίνει άπειρη διακύμανση και επομένως πολύ χονδρές ουρές. Παρόλο που υπάρχει αυτός ο περιορισμός της διακύμανσης, οι κατανομές της τάξης A είναι μεγάλης σημασίας, λόγω του ότι αποτελούν μια γενίκευση του Κεντρικού Οριακού Θεωρήματος (Gnedenko, 1943). Το Κεντρικό Οριακό Θεώρημα, που αναφέρει ότι το άθροισμα μεγάλου αριθμού ανεξάρτητων και ισόνομων μεταβλητών από μια κατανομή με πεπερασμένη διακύμανση θα τείνει να είναι κανονικά κατανομημένο, δε μπορεί να εφαρμοστεί για κάποιες μεταβλητές με άπειρη διακύμανση (π.χ. μεταβλητές που αντιπροσωπεύουν φυσικές καταστροφές). Για το λόγο αυτό, ο Fama (1965) και ο Mandelbrot (2003) πρότειναν τη σταθερή κατανομή σαν εναλλακτικό μοντέλο. Οι σταθερές κατανομές υποστηρίζονται από το γενικευμένο Κεντρικό Οριακό Θεώρημα, που αναφέρει ότι οι σταθεροί νόμοι είναι οι μόνες πιθανές οριακές κατανομές για ορθώς κανονικοποιημένα αθροίσματα ανεξάρτητων και ισόνομων τυχαίων μεταβλητών. Εφόσον οι σταθερές κατανομές μπορούν να «φιλοξενήσουν» τις χονδρές ουρές και την ασυμμετρία, συχνά ταιριάζουν πολύ καλά στα εμπειρικά δεδομένα. Συγκεκριμένα, είναι χρήσιμα μοντέλα για ακραία γεγονότα.

2.7.3 Κατά Embrechts et al.

(Embrechts et al., 1997)

Ένας άλλος τρόπος κατάταξης κατανομών βάσει της ουράς τους (Embrechts et al., 1997) προτείνει την ύπαρξη πάλι πέντε κατηγοριών κατανομών (Σχήμα 3.2):

- Την γενική τάξη K όπου

$$K = \left\{ F \text{ σ.κ. στο } (0, \infty) : \hat{f}(-\varepsilon) = \int_0^{\infty} e^{\varepsilon x} dF(x) = \infty \text{ για όλα τα } \varepsilon > 0 \right\}$$

- Την τάξη των βραδείας κύμανσης κατανομών L (των οποίων οι ουρές χαρακτηρίζονται ως μακριές) όπου:

$$L = \left\{ F \text{ σ.κ. στο } (0, \infty) : \lim_{x \rightarrow \infty} \bar{F}(x-y)/\bar{F}(x) = 1 \text{ για όλα τα } y > 0 \right\}$$

- Την τάξη των υποεκθετικών κατανομών S (ταχείας κύμανσης), που ταυτίζεται με την τάξη D της κατάταξης των Werner and Uppner (2002):

$$S = \left\{ F \text{ σ.κ. στο } (0, \infty) : \lim_{x \rightarrow \infty} \frac{\bar{F}^{n^*}(x)}{\bar{F}(x)} = n \text{ για όλα τα } n \geq 2 \right\}$$

όπου $F^{n^*}(x)$ δηλώνει τη n -αναδιπλούμενη συνέλιξη της F .

- Την τάξη D των κυρίως μεταβαλλόμενων κατανομών όπου

$$D = \left\{ F \text{ σ.κ. στο } (0, \infty) : \limsup_{x \rightarrow \infty} \bar{F}(x/2)/\bar{F}(x) < \infty \right\}$$

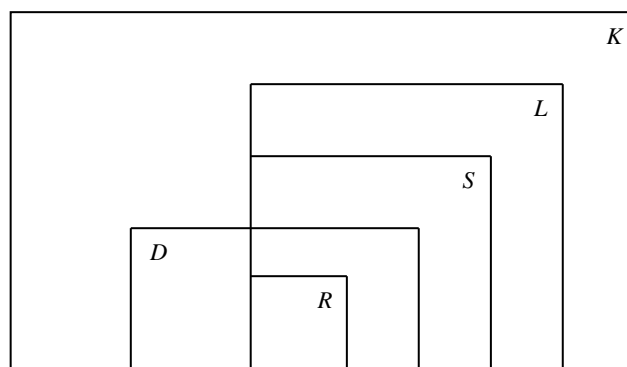
$$\text{ή } \limsup_{x \rightarrow \infty} \frac{\bar{F}(xy)}{\bar{F}(x)} < \infty \text{ για κάθε } y \in (0, 1) \text{ (ισοδύναμα για } y = \frac{1}{2} \text{)}$$

- Την τάξη των κατανομών ομαλής κύμανσης R όπου

$$R = \left\{ F \text{ σ.κ. στο } (0, \infty) : \bar{F} \in R_{-a} \text{ για κάποια } a \geq 0 \right\}$$

όπου $h \in R_a$ εάν $\lim_{x \rightarrow \infty} \frac{h(tx)}{h(x)} = t^a, t > 0$ (βλ.2.11)

$$\text{ή } \lim_{x \rightarrow \infty} \frac{\bar{F}(xy)}{\bar{F}(x)} = y^{-a} \text{ για κάποια } a \geq 0 \text{ και κάθε } y > 0.$$



Σχήμα 2.8 Ενδεικτικό σχήμα κατάταξης βαριών ουρών κατά Embrechts et al. (1997).). Προέλευση: Embrechts et al. (1997).

Όπως φαίνεται και από το σχήμα ισχύουν οι εξής ιδιότητες:

- (1) $R \subset S \subset L \subset K$ και $R \subset D$
- (2) $L \cap D \subset S$
- (3) $D \not\subset S$ και $S \not\subset D$

Οι Cline και Samorodnitsky (1994) προτείνουν μια παρόμοια κατάταξη, με τη διαφορά ότι δεν αναφέρουν την τάξη K , αλλά παρεμβάλλουν ανάμεσα στις R και D την τάξη C η οποία περιέχει κατανομές με «συνεπή κύμανση», για την οποία ισχύει ότι μια σ.κ. F ανήκει στη C αν

$$\lim_{y \uparrow 1} \limsup_{x \rightarrow \infty} \frac{\bar{F}(xy)}{\bar{F}(x)} = 1 \quad (2.27)$$

ή, ισοδύναμα,

$$\lim_{y \downarrow 1} \liminf_{x \rightarrow \infty} \frac{\bar{F}(xy)}{\bar{F}(x)} = 1 \quad (2.28)$$

Οι τάξεις σ' αυτήν την περίπτωση ακολουθούν την ακόλουθη σχέση «ενσωμάτωσης»:
 $R \subset C \subset D \cap L \subset S \subset L$.

2.7.4 Κατά Ouarda et al.

Οι Ouarda et al. (1994) παρουσίασαν μια κατηγοριοποίηση κατανομών που συνήθως χρησιμοποιείται στην υδρολογία και βασίζεται στην ασυμπτωτική συμπεριφορά της κατανομής πυκνότητας πιθανότητας. Για δεδομένη περίοδο επαναφοράς T αρκετά μεγάλη, ισχύει η επόμενη ισοδυναμία (Gumbel, 1958; Ouarda et al., 1994)

$$T \underset{x_T \rightarrow \infty}{\sim} \left(\frac{1}{f(x_T)} \right)' = -\frac{f'(x_T)}{f^2(x_T)} \quad (2.29)$$

Πράγματι, για μεγάλη περίοδο επαναφοράς T , η συνάρτηση πιθανότητας υπέρβασης $\bar{F}(x_T) = 1 - F(x_T) = \frac{1}{T}$ και η συνάρτηση πυκνότητας πιθανότητας τείνουν στο μηδέν. Με

χρήση του κανόνα de L'Hospital για μεγάλες T έχουμε: $\lim_{x_T \rightarrow \infty} \frac{f(x_T)}{1 - F(x_T)} \simeq \lim_{x \rightarrow \infty} \frac{f'(x_T)}{-f(x_T)}$

$$\Rightarrow \frac{1}{1 - F(x_T)} \underset{x_T \rightarrow \infty}{\sim} -\frac{f'(x_T)}{f^2(x_T)} = \left(\frac{1}{f(x_T)} \right)'$$

Χρησιμοποιώντας αυτή την ισοδυναμία, η ασυμπτωτική συμπεριφορά των κατανομών που χρησιμοποιούνται συχνά στην υδρολογία παρουσιάζεται στον Πίνακα 2.1. Από τον πίνακα αυτόν προκύπτουν τέσσερις ομάδες με βάση αυτή την κατάταξη.

- Η πρώτη περιέχει κατανομές των οποίων η ουρά είναι συνάρτηση δύναμης της περιόδου επαναφοράς ($x \approx T^p$).
- Η δεύτερη περιέχει τη Λογαριθμοκανονική κατανομή, της οποίας η ουρά έχει συμπεριφορά παρόμοια με αυτήν της κατανομής τύπου δύναμης.
- Η ουρά των κατανομών της τρίτης ομάδας είναι δυναμική συνάρτηση του λογαρίθμου της περιόδου επαναφοράς ($x \approx (\ln T)^p$) και όλες σχεδόν οι κατανομές αυτής της ομάδας ανήκουν στην τάξη D των υποεκθετικών κατανομών.
- Η τελευταία ομάδα περιέχει κατανομές άνω φραγμένες (upper bounded support).

Η κατάταξη αυτή από Ouarda et al. (1994) είναι ισοδύναμη με την προαναφερθείσα (Werner and Upper, 2002). Παρατηρούμε ότι οι κατανομές που ανήκουν στην πρώτη ομάδα του Πίνακα 2.1 έχουν δυναμικού τύπου ουρά και αντιστοιχούν γενικά στην τάξη C των

κατανομών ομαλής κύμανσης (όπου περιλαμβάνεται και η γενικευμένη Pareto με $\alpha > 0$ και $k < 0$). Όμως, για ειδικές τιμές παραμέτρων, αυτές οι κατανομές της τάξης C ανήκουν στην ομάδα των υποεκθετικών κατανομών (ομάδα III, Πίνακας 2.1). Εφόσον οι τάξεις περιέχονται η μία μέσα στην άλλη, η τάξη των κατανομών που ανήκουν στην τάξη C_2 και όχι στην τάξη C_1 και είναι $C_1 \subset C_2$ θα συμβολίζονται $C_2 \setminus C_1$. Για παράδειγμα, η τάξη των υποεκθετικών κατανομών που δεν είναι ομαλής κύμανσης συμβολίζεται . Συνδυάζοντας αυτές τις δύο κατατάξεις, οι κατανομές που χρησιμοποιούνται στην υδρολογία μπορούν να διαταχθούν σύμφωνα με τις ουρές τους (Πίνακας 2.2).

Πίνακας 2.1 Ταξινόμηση βάσει ασυμπτωτικής συμπεριφοράς συχνά χρησιμοποιούμενων στην υδρολογία κατανομών. Προέλευση: *Ouarda et al., 1994*

Ομάδα	Χαρακτηριστικά	Κατανομή	Παράμετροι
II: $x \approx T^P$			
Τάξη C	$P = 1/\alpha$	Log-Pearson 3 (α, λ, m)	$\alpha > 0, \lambda > 0, m \in \mathbb{R}$
Τάξη C	$P = 1/\alpha$	Log-logistic (α, λ)	$\alpha > 0, \lambda > 0$
Τάξη C	$P = -1/s\lambda$	Γενικευμένη Γάμμα (s, α, λ)	$s < 0, \alpha > 0, \lambda > 0$
Τάξη C	$P = 1/\lambda$	Αντίστροφη Γάμμα (α, λ)	$\alpha > 0, \lambda > 0$
Τάξη C	$P = -1/k$	Fréchet (α, k, u)	$\alpha > 0, k > 0, u \in \mathbb{R}$
Τάξη B	$P = -1/k$	Γενικευμένη Pareto (α, k)	$\alpha > 0, k \in \mathbb{R}$
Τάξη C	$1/2\nu$	Halphen B ⁻¹ (α, ν, m)	$\alpha \in \mathbb{R}, \nu > 0, m > 0$
II: $x \approx \exp\left[(\ln T)^{1/2}\right]$			
-		Λογαριθμοκανονική 2 (μ, σ)	$\sigma > 0, \mu \in \mathbb{R}$
-		Λογαριθμοκανονική 3 (μ, σ, m)	$m \in \mathbb{R}, \sigma > 0, \mu \in \mathbb{R}$
III: $x \approx (\ln T)^P$			
Τάξη D	$P = 1$	Pearson τύπου 3 (α, λ, m)	$\alpha > 0, \lambda > 0, m \in \mathbb{R}$
Τάξη D		Γάμμα ¹ (α, λ)	$\alpha > 0, \lambda > 0$
Τάξη E		Εκθετική (α, m)	$\alpha > 0, m \in \mathbb{R}$
Τάξη D		Halphen A (α, ν, m)	$\alpha > 0, \nu \in \mathbb{R}, m > 0$
Τάξη D		Log-F (λ, β)	$\beta > 0, \lambda > 0$
Τάξη D		Gumbel (α, u)	$\alpha > 0, u \in \mathbb{R}$
Τάξη D	$P = 1/2$	Halphen B (α, ν, m)	$\alpha \in \mathbb{R}, \nu > 0, m > 0$
Τάξη D	$P = 1/s$	Γενικευμένη Γάμμα (s, α, λ)	$s > 0, \alpha > 0, \lambda > 0$
IV: $x \leq P$			
(Ανω φραγμένες)	$P = m$	Pearson 3 (α, λ, m)	$\alpha < 0, \lambda > 0, m \in \mathbb{R}$
	$P = 0$	Γάμμα (α, λ)	$\alpha < 0, \lambda > 0$
	$P = \exp(m / \ln_\alpha(e))$	Log-Pearson 3 (β, γ, κ)	$\beta < 0, \gamma > 0, \kappa \in \mathbb{R}$
	$P = 1$	Log-Logistic (α, λ)	$\beta < 0, \gamma > 0$
	$P = u + \alpha/k$	Αντίστροφη Weibull 3	$\alpha > 0, k < 0, u \in \mathbb{R}$
	$P = \alpha/k$	Γενικευμένη Pareto	$\alpha > 0, k > 0$

¹ Η κατανομή Γάμμα σύμφωνα με τους Embrechts et al. (2003) ανήκει στην τάξη $S(\gamma)$ με $\gamma \geq 0$. Μια κατανομή F ανήκει στην τάξη αυτή όταν ισχύουν:

$$(a) \lim_{x \rightarrow \infty} \frac{\bar{F}^{2*}(x)}{\bar{F}(x)} = 2d < \infty \quad (b) \lim_{x \rightarrow \infty} \frac{\bar{F}(x-y)}{\bar{F}(x)} = e^{\gamma y}, y \in \mathbb{R}.$$

Πίνακας 2.2 Ταξινόμηση κατανομών συχνά χρησιμοποιούμενων στην Υδρολογία βάσει της συμπεριφοράς της ουράς.

Ταξινόμηση	Κατανομές
«Λεπτές» Ουρές	Κανονική
Τάξη $E \setminus D$	Εκθετική
Τάξη $D \setminus C$	Gumbel, Halphen τύπου A, Γάμμα, Halphen τύπου B, Pearson τύπου B
Ομάδα II	Λογαριθμοκανονική
Τάξη $C \setminus B$	Fréchet, Halphen B, Αντίστροφη Γάμμα, Λογαριθμική Pearson τύπου 3
Τάξη $B \setminus A$	Pareto
Τάξη A	Σταθερές Κατανομές

Η «βαρύτητα» της ουράς στον Πίνακα 2.2 αυξάνεται από πάνω προς τα κάτω.

3 ΑΝΑΛΥΣΗ

3.1 ΕΙΣΑΓΩΓΗ

3.1.1 Στατιστική Προσέγγιση

Η κατανομή της βροχόπτωσης ανήκει στο λεγόμενο ανάμικτο τύπο κατανομών, με ένα διακριτό μέρος που περιγράφει την πιθανότητα μηδενικής βροχόπτωσης ή την πιθανότητα ξηρασίας, και ένα συνεχές μέρος που εκφράζει το μέγεθος της μη μηδενικής βροχόπτωσης. Όπως έχουμε ήδη αναφέρει, στη μελέτη των ακραίων βροχοπτώσεων μας ενδιαφέρει η συμπεριφορά της δεξιάς ουράς της κατανομής που ελέγχει τη συχνότητα και το μέγεθος των ακραίων.

Αν θεωρήσουμε τη βροχόπτωση ως μεταβλητή X και τη μη μηδενική βροχόπτωση με $X | X > 0$, τότε η συνάρτηση υπέρβασης πιθανότητας της μη μηδενικής βροχόπτωσης ορίζεται ως

$$P(X > x | X > 0) = \bar{F}_{X|X>0}(x) = 1 - F_{X|X>0}(x) \quad (3.1)$$

όπου $F_{X|X>0}(x)$ είναι κάθε κατανομή πιθανότητας που έχει επιλεγεί για να περιγράψει μη μηδενική βροχόπτωση. Η Σχέση 3.1 αποτελεί τη συνάρτηση ουράς της πιθανοτικής κατανομής της βροχόπτωσης.

3.1.2 Ορισμός Ανώτερου Μέρους Κατανομής

Για να οριστεί το άνω μέρος της κατανομής, μια συνήθης πρακτική είναι η θεώρηση ενός κατώτατου ορίου (Cunnane, 1973; Tavares & Da Silva, 1983) και η μελέτη της συμπεριφοράς για τιμές μεγαλύτερες του x_L (Μέθοδος Peaks over Threshold). Δεν υπάρχει όμως μια γενικά αποδεκτή μέθοδος καθορισμού της κατώτατης τιμής. Το u πρέπει να επιλέγεται ικανοποιητικά μεγάλο. Όμως, μια πολύ μεγάλη τιμή του u έχει ως αποτέλεσμα πολύ λίγες υπερβάσεις, επομένως υψηλές εκτιμήτριες διακύμανσης. Για u πολύ μικρό από την άλλη, οι εκτιμήτριες γίνονται μεροληπτικές.

Εδώ θα γίνει χρήση μιας κοινώς αποδεκτής μεθόδου γνωστή ως Partial Duration Series Method. Με αυτή, το όριο ορίζεται έμμεσα με βάση την εμπειρική κατανομή έτσι ώστε ο

αριθμός των τιμών πάνω από το όριο να είναι ίσος με τον αριθμό των ετών των δεδομένων (Cunane, 1973).

Σε σύγκριση με την άλλη κοινή μέθοδο, όπου μελετώνται τα ετήσια μέγιστα, η μέθοδος που υιοθετείται εκπροσωπεί καλύτερα την ακριβή ουρά της μητρικής κατανομής. Η μέθοδος των ετήσιων μεγίστων, με το να επιλέγει τη μέγιστη τιμή κάθε χρόνου μπορεί να διαστρεβλώσει τη συμπεριφορά της ουράς (π.χ. όταν οι τρεις μεγαλύτερες ημερήσιες τιμές συμβαίνουν σε ένα χρόνο, η μέθοδος των ετήσιων μεγίστων λαμβάνει υπόψη μόνο τη μεγαλύτερη από αυτές). Για το λόγο αυτό, αντί να μελετηθούν τα N ημερήσια ετήσια μέγιστα, μελετούνται οι N μεγαλύτερες ημερήσιες τιμές και θεωρούνται αυτές ως αντιπροσωπευτικές για την ουρά.

3.1.3 Σχεδιαστικές Θέσεις

Τα δεδομένα, που μπορεί να είναι είτε τα ετήσια μέγιστα είτε οι τιμές πάνω από ένα κατώφλι, ταξινομούνται σε αύξουσα σειρά. Η μικρότερη παρατήρηση δηλώνεται με X_1 και η μεγαλύτερη με X_n . Στη συνέχεια για κάθε παρατήρηση υπολογίζεται η εκτίμηση της εμπειρικής πιθανότητας μη υπέρβασης p_i που αναφέρεται και ως σχεδιαστική θέση. Αυτή δίνεται από την παρακάτω γενική σχέση:

$$p_i = \frac{i - a}{n + 1 - 2a} \quad (3.2)$$

όπου p_i η πιθανότητα μιας τιμής να είναι μικρότερη από την i -οστή μικρότερη παρατήρηση σε ένα δείγμα n παρατηρήσεων. Η παράμετρος a είναι μια σταθερά που παίρνει τιμές μεταξύ 0 και 0.5. Στην εκτίμηση των περιόδων επαναφοράς η πιο σωστή θέση σχεδίασης είναι η:

$$p_i = \frac{i}{n + 1} \quad (3.3)$$

Αυτή η σχέση αποδεικνύεται ότι προβλέπει πολύ πιο σύντομες περιόδους επαναφοράς ακραίων γεγονότων σε σχέση με τις άλλες χρησιμοποιούμενες (Makkonen 2006). Είναι γνωστή ως σχεδιαστική θέση κατά Weibull.

3.2 ΥΠΟ ΕΞΕΤΑΣΗ ΤΥΠΟΙ ΚΑΤΑΝΟΜΩΝ

3.2.1 Εκθετική Κατανομή

Η Εκθετική Κατανομή, η ουρά της οποίας χαρακτηρίζεται ως «ελαφριά», έχει παράμετρο κλίμακας αντιστρόφως ανάλογη της παραμέτρου β . Ανήκει στην ομάδα III ή στην τάξη $E \setminus D$ (βλ. Σχήμα 2.6)

Πίνακας 3.1 Τυπολόγιο Εκθετικής κατανομής

Χαρακτηριστικά	Εξίσωση
Περιορισμοί	$\beta > 0, x > 0$
Πυκνότητα Πιθανότητας	$f(x; \beta) = \frac{1}{\beta} \exp\left(-\frac{x}{\beta}\right)$
Συνάρτηση Κατανομής	$F(x; \beta) = 1 - \exp\left(-\frac{x}{\beta}\right)$
Ποσοστημόριο	$x_q = -\beta \log(1-q), 0 \leq q < 1$
Ουρά	$\bar{F}(x; \beta) = \exp\left(-\frac{x}{\beta}\right)$

3.2.2 Κατανομή Pareto τύπου II

Η διπαραμετρική κατανομή Pareto τύπου II έχει παράμετρο κλίμακας β και σχήματος γ (βλ. Σχήμα 2.7). Ανήκει στην ομάδα I, στην τάξη $B \setminus A$.

Πίνακας 3.2 Τυπολόγιο κατανομής Pareto

Χαρακτηριστικά	Εξίσωση
Περιορισμοί	$\beta, \gamma > 0, x > 0$
Πυκνότητα Πιθανότητας	$f(x; \beta, \gamma) = \frac{1}{\beta} \left(1 + \gamma \frac{x}{\beta}\right)^{-\frac{1}{\gamma}}$
Συνάρτηση Κατανομής	$F(x; \beta, \gamma) = 1 - \left(1 + \gamma \frac{x}{\beta}\right)^{-\frac{1}{\gamma}}$
Ποσοστημόριο	$\frac{\beta \left((1-q)^{-\gamma} - 1\right)}{\gamma}$
Ουρά	$\bar{F}(x; \beta, \gamma) = \left(1 + \gamma \frac{x}{\beta}\right)^{-\frac{1}{\gamma}}$

3.2.3 Λογαριθμοκανονική Κατανομή

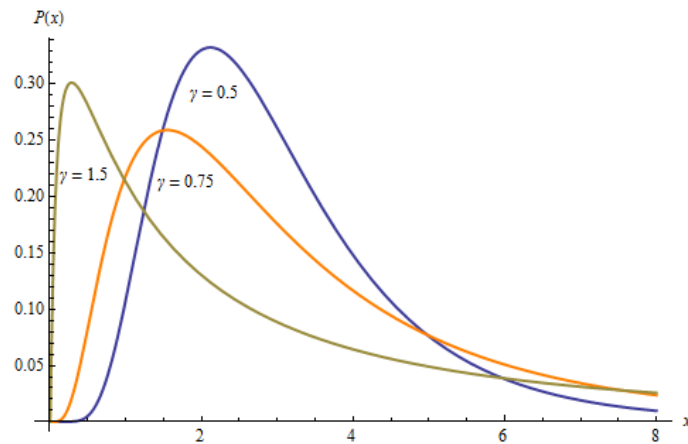
Η Λογαριθμοκανονική κατανομή προέρχεται από μια Κανονική Κατανομή με μέση τιμή μ και τυπική απόκλιση σ (εάν η μεταβλητή x ακολουθεί Κανονική Κατανομή, τότε η ποσότητα $\exp(x)$ ακολουθεί Λογαριθμοκανονική κατανομή). Εμφανίζεται πολύ συχνά στη στοχαστική χρηματοοικονομική ανάλυση και ανήκει στην ομάδα II (Πίνακες 2.1, 2.2). Εδώ, έχοντας θεωρήσει την παράμετρο θέσης ίση με μηδέν, η κατανομή είναι διπαραμετρική και συμβολίζεται με β η παράμετρος κλίμακας που είναι η μέση τιμή και με γ η παράμετρος σχήματος που είναι η τυπική απόκλιση (Σχήμα 3.1).

Πίνακας 3.3 Τυπολόγιο Λογαριθμοκανονικής κατανομής

Χαρακτηριστικά	Εξίσωση
Περιορισμοί	$\beta, \gamma > 0, x > 0$
Πυκνότητα Πιθανότητας	$f(x; \beta, \gamma) = \frac{\exp\left(-\frac{(\ln x - \beta)^2}{2\gamma^2}\right)}{\sqrt{2\pi}\gamma x}$
Συνάρτηση Κατανομής	$F(x; \mu, \sigma) = \frac{1}{2} + \frac{1}{2} \operatorname{erfc}\left(\frac{\ln x - \beta}{\sqrt{2}\gamma}\right)$
Ποσοστημόριο	$x_q = \exp\left(\beta + \sqrt{2}\gamma \operatorname{erf}^{-1}(-1 + 2q)\right)$
Ουρά	$\bar{F}(x; \beta, \gamma) = \frac{1}{2} \operatorname{erfc}\left(\frac{\ln x - \beta}{\sqrt{2}\gamma}\right)$

Όπου erfc η συμπληρωματική συνάρτηση σφάλματος που ορίζεται ως

$$\operatorname{erfc}(z) = 1 - \operatorname{erf}(z) = 1 - \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt .$$



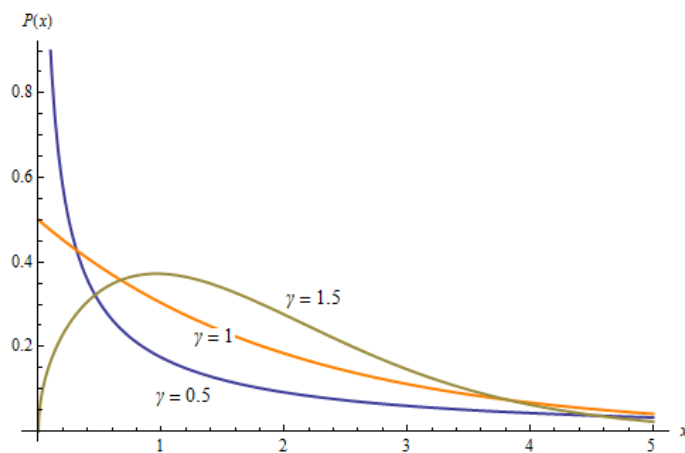
Σχήμα 3.1 Συνάρτηση πυκνότητας πιθανότητας Λογαριθμοκανονικής κατανομής για διάφορες τιμές της παραμέτρου σχήματος γ . Η παράμετρος κλίμακας είναι $\beta = 1$.

3.2.4 Κατανομή Weibull

Η κατανομή Weibull μπορεί και αυτή να θεωρηθεί γενίκευση της Εκθετικής κατανομής. Είναι διπαραμετρική με παράμετρο σχήματος γ και παράμετρο κλίμακας β (Σχήμα 3.2).

Πίνακας 3.4 Τυπολόγιο κατανομής Weibull

Χαρακτηριστικά	Εξίσωση
Περιορισμοί	$\gamma, \beta > 0, x > 0$
Πυκνότητα Πιθανότητας	$f(x; \gamma, \beta) = \frac{\gamma x^{\gamma-1} \exp\left[-(x/\beta)^\gamma\right]}{\beta^\gamma}$
Συνάρτηση Κατανομής	$F(x; \gamma, \beta) = 1 - \exp\left[-\left(\frac{x}{\beta}\right)^\gamma\right]$
Ποσοστημόριο	$x_q = \beta(-\log(1-q))^{\frac{1}{\gamma}}, 0 < q < 1$
Ουρά	$\bar{F}(x; \gamma, \beta) = \exp\left[-\left(\frac{x}{\beta}\right)^\gamma\right]$



Σχήμα 3.2 Συνάρτηση πυκνότητας πιθανότητας κατανομής Weibull για διάφορες τιμές της παραμέτρου σχήματος γ . Η παράμετρος κλίμακας είναι $\beta = 3$.

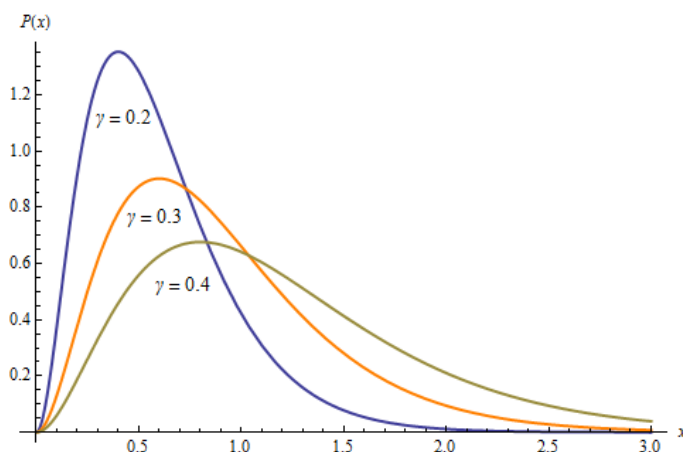
3.2.5 Κατανομή Γάμμα

Η κατανομή Γάμμα είναι γενίκευση της Εκθετικής κατανομής και ανήκει στην ομάδα II ή στην τάξη $D \setminus C$ (σύμφωνα με τους Embrechts et al. (2003) ανήκει στην τάξη $S(\gamma)$ βλ. Πίνακες 3.1 και 3.2). Είναι διπαραμετρική με παράμετρο σχήματος γ και παράμετρο κλίμακας β (Σχήμα 3.3).

Πίνακας 3.5 Τυπολόγιο κατανομής Γάμμα

Χαρακτηριστικά	Εξίσωση
Περιορισμοί	$\gamma, \beta > 0, x > 0$
Πυκνότητα Πιθανότητας	$f(x; \gamma, \beta) = \frac{\exp(-x/\beta) x^{\gamma-1} \beta^{-\gamma}}{\Gamma(\gamma)}$
Συνάρτηση Κατανομής	$F(x; \gamma, \beta) = \int_0^x \frac{\exp(-t/\beta) t^{\gamma-1} \beta^{-\gamma}}{\Gamma(\gamma)} dt$
Ποσοστημόριο	Δεν έχει αναλυτική έκφραση
Ουρά	$\bar{F}(x; \gamma, \beta) = 1 - \int_0^x \frac{\exp(-t/\beta) t^{\gamma-1} \beta^{-\gamma}}{\Gamma(\gamma)} dt$

όπου $\Gamma(\gamma) = \int_0^\infty t^{\gamma-1} \exp(-t) dt$ και $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$, $\Gamma(n) = \Gamma(n-1)!$, $n \in \mathbb{N}$, $\Gamma(1/2) = \sqrt{\pi}$



Σχήμα 3.3 Συνάρτηση πυκνότητας πιθανότητας κατανομής Γάμμα για διάφορες τιμές της παραμέτρου σχήματος γ . Η παράμετρος κλίμακας είναι $\beta = 2$.

3.3 ΥΠΟ ΕΞΕΤΑΣΗ ΟΥΡΕΣ ΚΑΤΑΝΟΜΩΝ

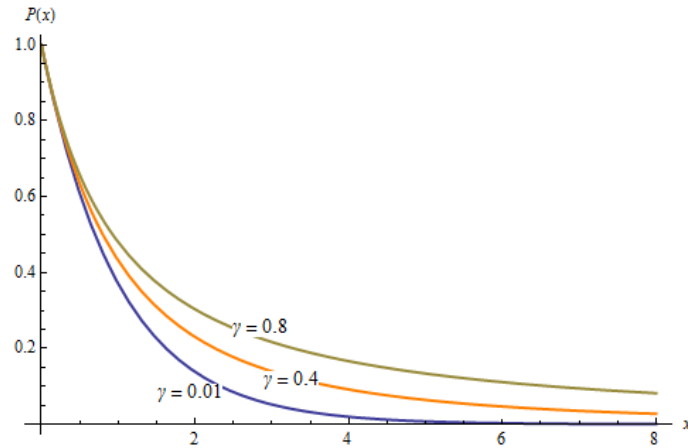
Θα μελετηθεί η συμπεριφορά τεσσάρων διαφορετικών συχνά χρησιμοποιούμενων τύπων ουρών: της Pareto τύπου II, της Λογαριθμοκανονικής, της Weibull και της Γάμμα, των οποίων οι ασυμπτωτική συμπεριφορά μπορεί να είναι παρόμοια με πολλών άλλων κατανομών. Συγκεκριμένα, επιλέχθηκε η Γάμμα ως συχνά χρησιμοποιούμενη, η Weibull ως απλή γενίκευση της Εκθετικής, η Pareto ως τύπου δύναμης και η Λογαριθμοκανονική ως ενδιάμεση. Οι κατανομές Pareto και Λογαριθμοκανονική ανήκουν στην τάξη των υποεκθετικών κατανομών, άρα θεωρούνται κατανομές με βαριές ουρές. Οι κατανομές Weibull και Γάμμα ανάλογα με τις τιμές της παραμέτρου τους σχήματος μπορεί να ανήκουν είτε στην τάξη των υποεκθετικών, είτε σε αυτή των υπερεκθετικών, σε κάθε περίπτωση όμως οι ουρές τους είναι πιο ελαφριές από αυτές των Pareto και Λογαριθμοκανονικής (από Papalexiou et al., 2012).

3.3.1 Ουρά Κατανομής Pareto τύπου II

Η κατανομή Pareto τύπου II είναι η απλούστερη κατανομή τύπου δύναμης που ορίζεται στο $[0, \infty)$. Η ουρά της δίνεται από τον τύπο

$$\bar{F}(x) = \left(1 + \gamma \frac{x}{\beta}\right)^{-\frac{1}{\gamma}} \quad (3.4)$$

και ορίζεται από την παράμετρο κλίμακας $\beta > 0$ και την παράμετρο σχήματος $\gamma \geq 0$ που ελέγχει την ασυμπτωτική συμπεριφορά της ουράς (Σχήμα 3.4). Συγκεκριμένα, όσο αυξάνεται η τιμή του γ , η ουρά γίνεται πιο βαριά, συνεπώς οι ακραίες τιμές θεωρείται ότι συμβαίνουν πιο συχνά. Για $\gamma = 0$ η ουρά εκφυλίζεται σε εκθετική ενώ για $\gamma \geq 0.5$ η κατανομή έχει άπειρη διακύμανση. Πολλές άλλες κατανομές δυναμικού τύπου έχουν ισοδύναμες ουρές, π.χ. παρουσιάζουν ασυμπτωτική συμπεριφορά παρόμοια με αυτή της κατανομής Pareto τύπου II (όπου έχουμε $x^{-1/\gamma}$). Μια τέτοια κατανομή είναι η Burr τύπου XII (Tadikamalla, 1980), η δι-και τρι-παραμετρική Κάππα (Mielke 1973), η Log-Logistic (Ahmad et al., 1988) και η γενικευμένη Βήτα του δεύτερου είδους (Mielke Jr & Johnson, 1974)



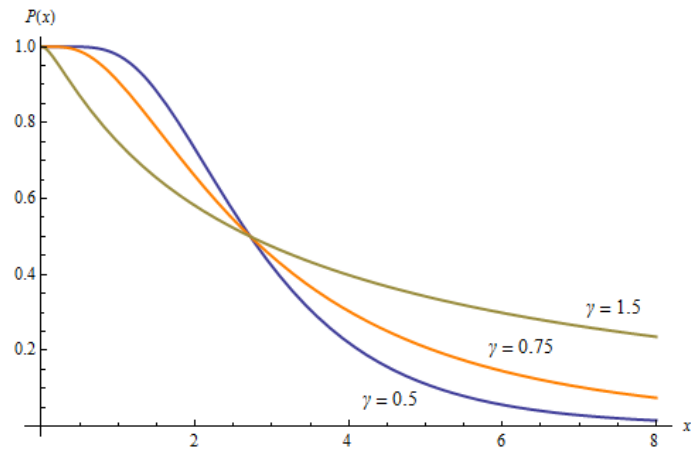
Σχήμα 3.4 Συνάρτηση ουράς κατανομής Pareto τύπου II για διάφορες τιμές της παραμέτρου σχήματος γ . Η παράμετρος κλίμακας είναι $\beta = 1$. Με την αύξηση της παραμέτρου σχήματος αυξάνεται το πάχος ουράς.

3.3.2 Ουρά Λογαριθμοκανονικής Κατανομής

Μια άλλη πολύ συνηθισμένη κατανομή που χρησιμοποιείται στην υδρολογία είναι η Λογαριθμοκανονική με συνάρτηση ουράς

$$\bar{F}(x) = \frac{1}{2} \operatorname{erfc}\left(\frac{\ln x - \beta}{\sqrt{2}\gamma}\right) \quad (3.5)$$

όπου $\operatorname{erfc}(x) = 2\pi^{-1/2} \int_x^{\infty} e^{-t^2} dt$. Η κατανομή περιλαμβάνει την παράμετρο κλίμακας $\beta \in \mathbb{R}$ και την παράμετρο $\gamma > 0$ που ελέγχει το σχήμα και τη συμπεριφορά της ουράς (Σχήμα 3.5). Η Λογαριθμοκανονική κατανομή επίσης έχει βαριά ουρά (ανήκει στην οικογένεια των υποεκθετικών κατανομών) και θεωρείται ότι μπορεί να προσεγγίσει δυναμικού τύπου κατανομές για μεγάλο ποσοστό του σώματος της κατανομής (Mitzenmacher, 2003).



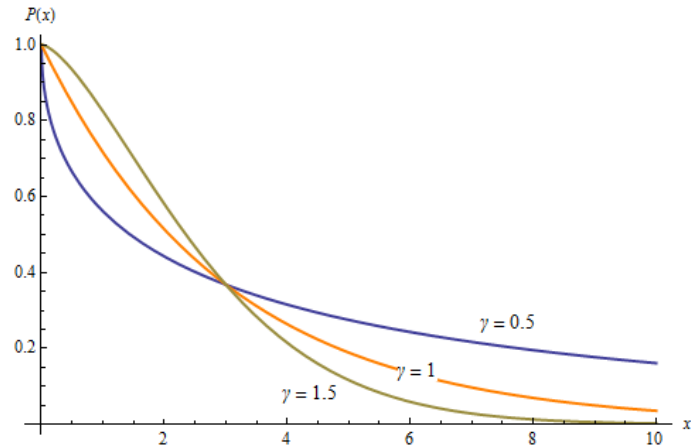
Σχήμα 3.5 Συνάρτηση ουράς Λογαριθμοκανονικής κατανομής για διάφορες τιμές της παραμέτρου σχήματος γ . Η παράμετρος κλίμακας είναι $\beta = 1$. Με την αύξηση της παραμέτρου σχήματος αυξάνεται το πάχος ουράς.

3.3.3 Ουρά Κατανομής Weibull

Η κατανομή Weibull, που θεωρείται γενίκευση της Εκθετικής Κατανομής είναι άλλο ένα σύνηθες μοντέλο στην υδρολογία (Heo, Boes, & Salas, 2001a) και η αντίστοιχη συνάρτηση ουράς είναι

$$\bar{F}(x) = \exp\left(-\frac{x^\gamma}{\beta}\right) \quad (3.6)$$

Η παράμετρος $\beta > 0$ είναι παράμετρος κλίμακας, ενώ η παράμετρος σχήματος $\gamma > 0$ ορίζει και πάλι την ασυμπτωτική συμπεριφορά της ουράς (Σχήμα 3.6). Για $\gamma < 1$ η κατανομή ανήκει στην οικογένεια των υποεκθετικών, με ουρά βαρύτερη από αυτήν της Εκθετικής, ενώ για $\gamma > 1$ η κατανομή χαρακτηρίζεται ως υπερεκθετική με ουρά λεπτότερη από της Εκθετικής. Πολλές κατανομές θεωρείται ότι έχουν ουρά ισοδύναμη με της Weibull για συγκεκριμένη τιμή της παραμέτρου γ , όπως π.χ. η Εκθετική, η Γενικευμένη Εκθετική, η Λογιστική και η Κανονική.



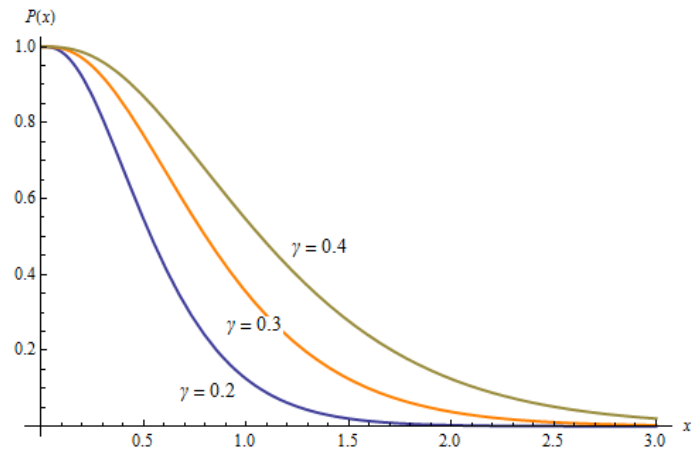
Σχήμα 3.6 Συνάρτηση ουράς κατανομής Weibull για διάφορες τιμές της παραμέτρου σχήματος γ . Η παράμετρος κλίμακας είναι $\beta = 3$.

3.3.4 Ουρά Κατανομής Γάμμα

Ένα από τα δημοφιλέστερα μοντέλα για την περιγραφή ημερήσιας βροχόπτωσης είναι η κατανομή Γάμμα (Buishand, 1978), που έχει λεπτή ουρά και ανήκει στην τάξη $S(\gamma)$. Η συνάρτηση ουράς της είναι

$$\bar{F}(x) = \Gamma\left(\gamma, \frac{x}{\beta}\right) / \Gamma(\gamma) \quad (3.7)$$

με $\Gamma(s, x) = \int_x^\infty t^{s-1} e^{-t} dt$ και $\Gamma(s) = \int_0^\infty t^{s-1} e^{-t} dt$ (Σχήμα 3.7). Γενικά μπορούμε να υποθέσουμε ότι η ουρά της κατανομής Γάμμα συμπεριφέρεται παρόμοια με αυτήν της Εκθετικής. Όμως, αυτό είναι μόνο κατά προσέγγιση σωστό, καθώς η εκθετική ουρά δεν είναι ασυμπτωτικά ισοδύναμη με την ουρά της Γάμμα, αφού η ασυμπτωτική συμπεριφορά της δεξιάς ουράς της δεύτερης εξαρτάται από την παράμετρο γ .

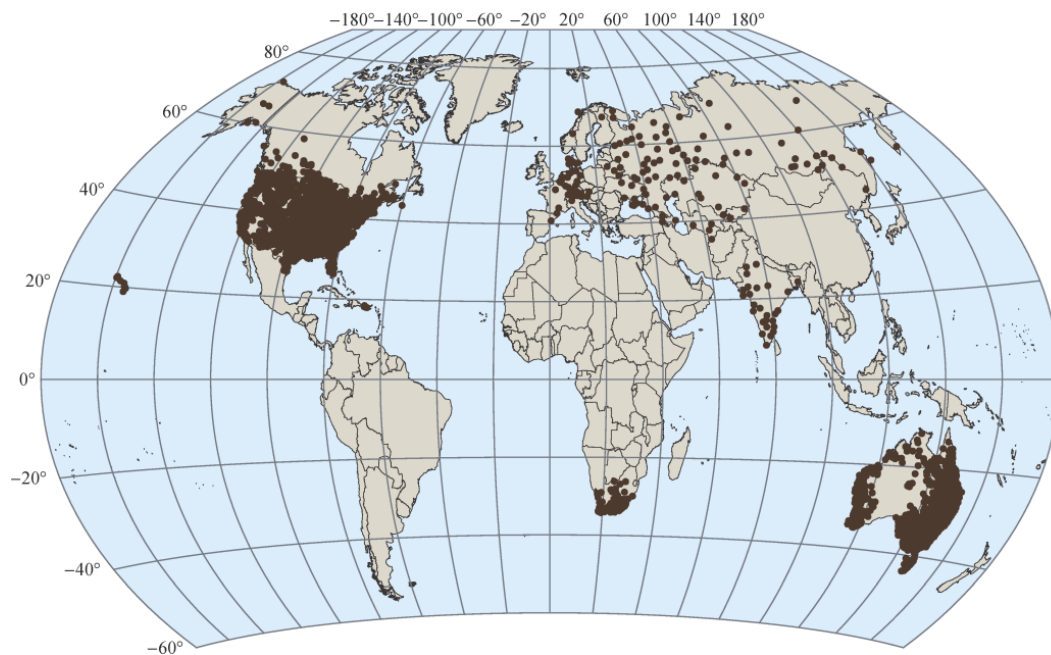


Σχήμα 3.7 Συνάρτηση ουράς κατανομής Γάμμα για διάφορες τιμές της παραμέτρου σχήματος γ . Η παράμετρος κλίμακας είναι $\beta = 2$.

4 ΜΕΘΟΔΟΛΟΓΙΑ

4.1 ΔΕΔΟΜΕΝΑ

Τα δεδομένα που μελετήθηκαν είναι οι ημερήσιες τιμές βροχόπτωσης από το Παγκόσμιο Ιστορικό Δίκτυο Κλιματολογίας - Ημερήσια Βάση δεδομένων (Global Historical Climatology Network, version 2.60, www.ncdc.noaa.gov/oa/climate/ghcn-daily) το οποίο περιέχει δεδομένα από 40 000 σταθμούς σε όλο τον κόσμο. Για την ανάλυση χρησιμοποιήθηκαν σταθμοί που είχαν καταγραφές 100 ετών και άνω, που δεν τους έλειπε πάνω από το 20% των δεδομένων για τα οποία η αμφισβητούμενη ποιότητα ήταν μικρότερη του 0.1%. Στις περιπτώσεις που υπήρχαν υπερβολικά μεγάλες τιμές δεδομένων, όταν π.χ. μια τιμή ήταν κάποιες τάξεις μεγέθους μεγαλύτερη από τη δεύτερη μεγαλύτερη του δείγματος, η πρώτη διαγραφόταν. Αυτό συνέβη σε 40 περίπου σταθμούς. Λαμβάνοντας υπόψη τα παραπάνω, τα δεδομένα προς ανάλυση προήλθαν από 3 477 σταθμούς. Στο Σχήμα 4.1 απεικονίζεται η γεωγραφική θέση των υπό εξέταση σταθμών.



Σχήμα 4.1 Γεωγραφική απεικόνιση των υπό εξέταση σταθμών

4.2 ΓΡΑΦΙΚΗ ΠΑΡΑΣΤΑΣΗ ΣΕ ΛΟΓΑΡΙΘΜΙΚΟΥΣ ΑΞΟΝΕΣ (LOG-LOG PLOT)

4.2.1 Θεωρητικό Υπόβαθρο

Η γραφική παράσταση σε λογαριθμικούς άξονες ίσως η πιο διαδεδομένη μέθοδος για τη μελέτη συμπεριφοράς ουρών (ιδιαίτερα στον τομέα των οικονομικών). Ουσιαστικά δημιουργείται γράφημα δύο διαστάσεων με λογαριθμικές κλίμακες και στους δύο άξονες. Στον άξονα των x είναι τα διατεταγμένα δεδομένα και στον άξονα των y το συμπληρωματικό της θέσης σχεδίασης, δηλαδή η πιθανότητα υπέρβασης. Εξαιτίας της μη γραμμικής κλίμακας των αξόνων, μια συνάρτηση της μορφής $y = ax^b$ θα εμφανίζεται σαν μια ευθεία γραμμή στο διπλό λογαριθμικό γράφημα ($\log y = \log a + b \log x$), στο οποίο το b θα είναι η κλίση της γραμμής και το a θα είναι η τιμή του y που αντιστοιχεί στο $x = 1$. Επειδή για μια εκθετική ουρά με μέση τιμή β είναι,

$$\bar{F}(u) = P(X > u) = \exp\left(-\frac{u}{\beta}\right) \quad (4.1)$$

ενώ για μια ουρά δύναμης με δείκτη ουράς $a > 1$, η \bar{F} είναι ισοδύναμη (για μεγάλα x) με:

$$\bar{F}(u) = P(X > u) \approx C \int_u^{\infty} \frac{1}{x^a} dx = C \frac{1}{a-1} u^{-a+1} \quad (4.2)$$

Έτσι, παίρνοντας το λογάριθμο, έχουμε για τις κατανομές εκθετικού τύπου:

$$\log[P(X > u)] \approx -u/\beta \quad (4.3)$$

και για δυναμικού τύπου κατανομές:

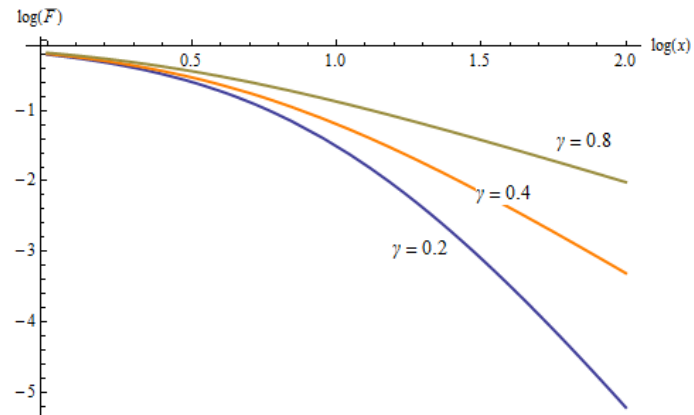
$$\log[P(X > u)] \approx \log\left(C \frac{1}{a-1}\right) - (1-a)\log(u) \quad (4.4)$$

Αυτό υποδηλώνει ότι για τη γραφική παράσταση σε λογαριθμικούς άξονες, η πιθανότητα ουράς αντιπροσωπεύεται από μια ευθεία γραμμή για δυναμικού τύπου κατανομές (ή ομαλής κύμανσης, τάξη C) αλλά όχι για τις άλλες υποεκθετικές ή εκθετικές κατανομές (τάξεις D/C ή E/D).

4.2.2 Εφαρμογή μεθόδου σε βασικές θεωρητικές κατανομές

Με εφαρμογή στις θεωρητικές κατανομές έχουμε (Σχήματα 4.2 έως 4.6):

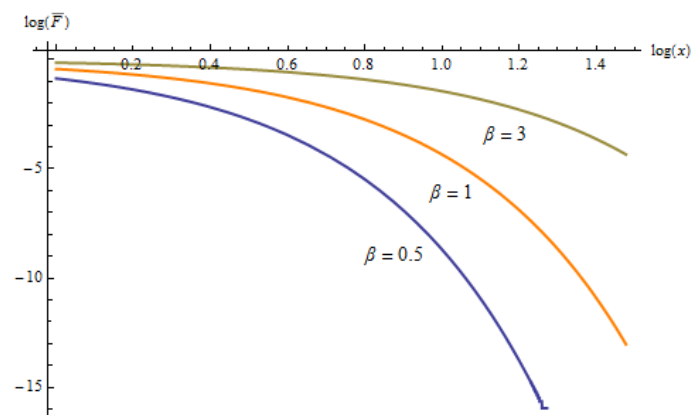
(α) Κατανομή Pareto τύπου II



Σχήμα 4.2 Συνάρτηση ουράς κατανομής Pareto σε διπλή λογαριθμική κλίμακα για διάφορες τιμές της παραμέτρου σχήματος γ . Η παράμετρος κλίμακας είναι $\beta = 2$.

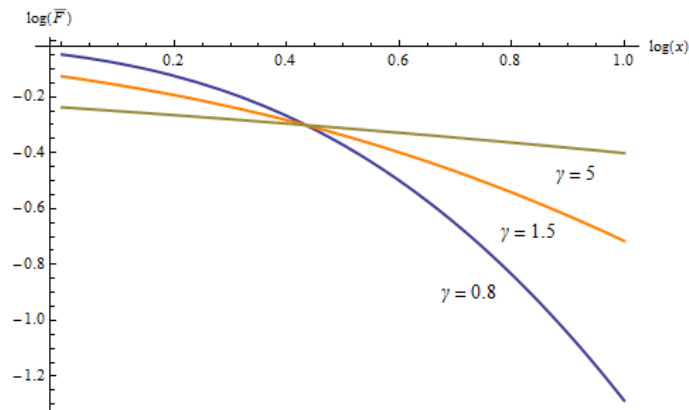
Όπως είναι φανερό (Σχήμα 4.2) η κατανομή Pareto εκφράζεται με ευθεία γραμμή στο διπλό λογαριθμικό διάγραμμα.

(β) Εκθετική Κατανομή



Σχήμα 4.3 Συνάρτηση ουράς Εκθετικής κατανομής σε διπλή λογαριθμική κλίμακα για διάφορες τιμές της παραμέτρου σχήματος β .

(γ) Λογαριθμοκανονική Κατανομή



Σχήμα 4.4 Συνάρτηση ουράς Λογαριθμοκανονικής κατανομής σε διπλή λογαριθμική κλίμακα για διάφορες τιμές της παραμέτρου σχήματος γ . Η παράμετρος κλίμακας είναι $\beta = 1$.

Παρατήρηση: Το συγκεκριμένο τεστ παρουσιάζει δυσκολία στο να διαχωρίσει τη Λογαριθμοκανονική ουρά από την ουρά Pareto. Και οι δύο κατανομές είναι ομαλά κυμαινόμενες, δηλαδή έχουν βαριές ουρές και οι ουρές τους έχουν παρόμοιες ιδιότητες. Συγκεκριμένα, για μεγάλη τυπική απόκλιση (δηλαδή μεγάλη τιμή της παραμέτρου γ) τόσο η Pareto όσο και η Λογαριθμοκανονική κατανομή έχουν γραμμικό λογάριθμο πυκνότητας (Mitzenmacher, 2003). Η Pareto τύπου II έχει λογάριθμο πυκνότητας (για μεγάλα x) που δίνεται από τον τύπο:

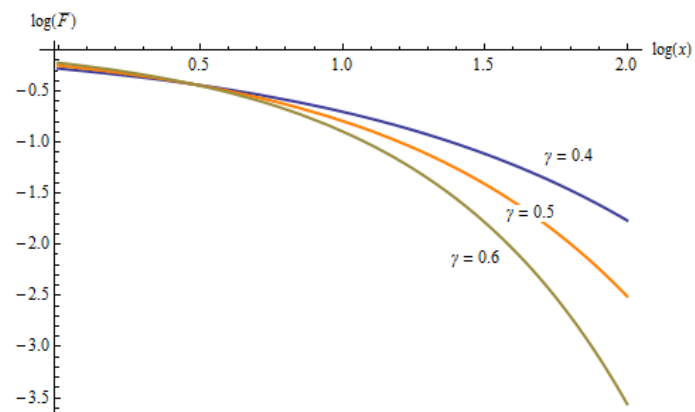
$$\ln f(x) = \ln \beta^{-1} - 1/\gamma \left(\ln \frac{\gamma}{\beta} + \ln x \right) \quad (4.5)$$

Ενώ η Λογαριθμοκανονική έχει λογάριθμο πυκνότητας

$$\ln f(x) = -\ln x - \ln(\sqrt{2\pi}\gamma) - \frac{(\ln x - \beta)^2}{2\gamma} \quad (4.6)$$

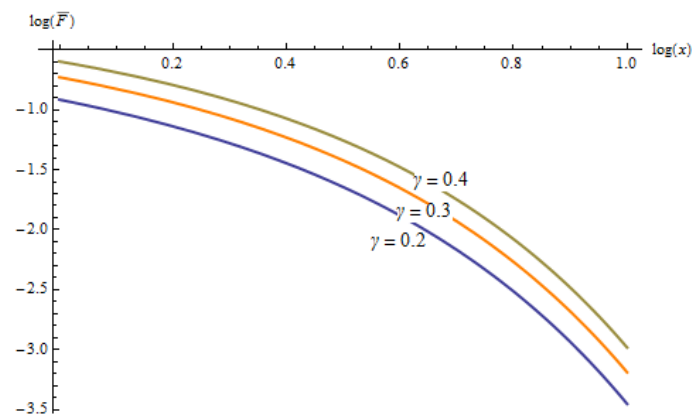
Ο λογάριθμος πυκνότητας της Pareto είναι γραμμική συνάρτηση στο $\ln x$ και το ίδιο ισχύει για την Λογαριθμοκανονική κατανομή για υψηλές τιμές της παραμέτρου γ , γεγονός που καθιστά το διαχωρισμό των ουρών των δύο κατανομών δυσδιάκριτο με αυτό το τεστ.

(δ) Κατανομή Weibull



Σχήμα 4.5 Συνάρτηση ουράς κατανομής Weibull σε διπλή λογαριθμική κλίμακα για διάφορες τιμές της παραμέτρου σχήματος γ . Η παράμετρος κλίμακας είναι $\beta = 3$.

(ε) Κατανομή Γάμμα



Σχήμα 4.6 Συνάρτηση ουράς κατανομής Γάμμα σε διπλή λογαριθμική κλίμακα για διάφορες τιμές της παραμέτρου σχήματος γ . Η παράμετρος κλίμακας είναι $\beta = 2$.

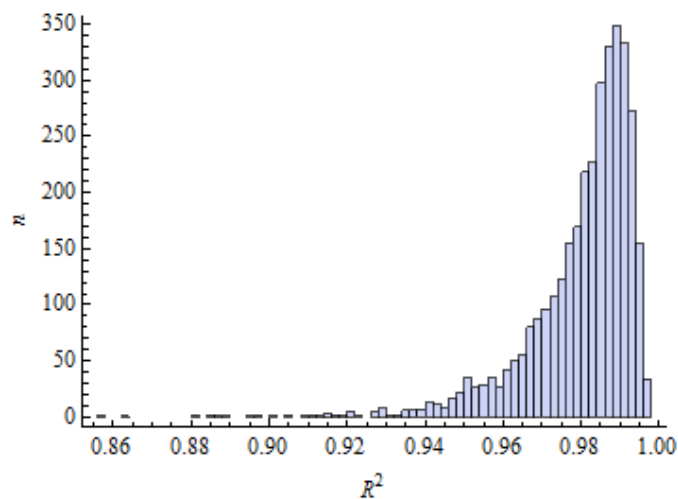
Στις περιπτώσεις των κατανομών Weibull και Γάμμα η κύρτωση του διαγράμματος είναι εμφανής (Σχήματα 4.5 και 4.6)

4.2.3 Εφαρμογή Μεθόδου στα δεδομένα

Για κάθε σταθμό εφαρμόστηκε η μέθοδος στις N υψηλότερες τιμές, όπου N ο αριθμός των ετών κάθε σταθμού. Συγκεκριμένα, αντιστοιχήθηκε ο λογάριθμος των διατεταγμένων σε αύξουσα σειρά δεδομένων με το λογάριθμο του συμπληρωματικού της σχεδιαστικής θέσης, ώστε να οριστεί η πιθανότητα υπέρβασης. Στο προκύπτον γράφημα, έγινε προσαρμογή ευθείας και βρέθηκε για αυτή η τιμή του συντελεστή προσδιορισμού R^2 , του γνωστού στατιστικού μέτρου που δείχνει πόσο καλά προσεγγίζονται τα πραγματικά δεδομένα από

ευθεία (ορίζεται ως $R^2 = 1 - \frac{SS_{\text{err}}}{SS_{\text{tot}}}$, όπου $SS_{\text{tot}} = \sum_i (y_i - \bar{y})^2$, $SS_{\text{err}} = \sum_i (y_i - \hat{y}_i)^2$).

Θεωρητικά, υψηλές τιμές του δείκτη R^2 υποδεικνύουν γραμμική σχέση και στην προκειμένη περίπτωση υποδηλώνουν ουρές τύπου δύναμης. Τα αποτελέσματα της εφαρμογής της παραπάνω διαδικασίας στα δεδομένα παρουσίασαν ενδιαφέρον, καθώς οι ουρές της πλειοψηφίας των σταθμών είχαν ιδιαίτερα υψηλές τιμές του συντελεστή R^2 , επομένως θα μπορούσαμε να οδηγηθούμε στο συμπέρασμα ότι το μεγαλύτερο μέρος των σταθμών έχουν ουρές Pareto ή (όπως αναφέρθηκε παραπάνω) Λογαριθμοκανονικές. Ακολουθεί ιστόγραμμα με τις τιμές του συντελεστή R^2 (Σχήμα 4.6). Συγκεκριμένα, ο μέσος όρος των τιμών του συντελεστή είναι 0.98 με τυπική απόκλιση 0.014 και η μέγιστη και ελάχιστη τιμή 0.997 και 0.857 αντίστοιχα.



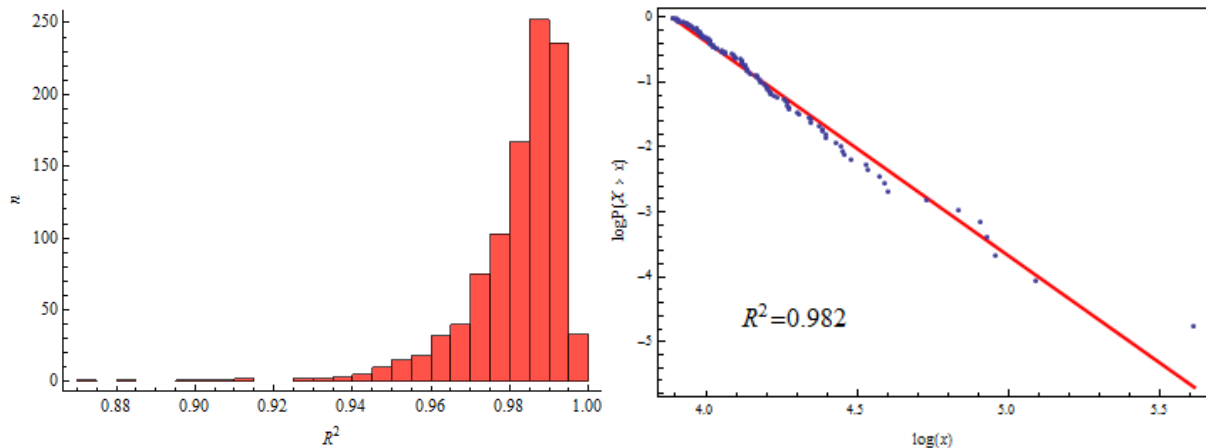
Σχήμα 4.7 Ιστόγραμμα τιμών συντελεστή R^2 για τα εμπειρικά δεδομένα

4.2.4 Προσομοίωση Monte Carlo

Πραγματοποιήθηκε επιπλέον και προσομοίωση Monte Carlo σε θεωρητικές κατανομές ώστε να ελεγχθεί η αξιοπιστία των αποτελεσμάτων. Παρήχθησαν 1.000 τυχαία δείγματα από τις κατανομές προς εξέταση και συγκρίθηκαν τα αποτελέσματα. Το μέγεθος του δείγματος λήφθηκε ίσο με το μέσο όρο των μετρήσεων $n = 9500$ και ο αριθμός δεδομένων ουράς ίσος με το μέσο όρο των ετών των μετρήσεων όλων των σταθμών, δηλαδή $N = 116$.

(α) Κατανομή Pareto με παράμετρο κλίμακας $\beta = 8$ και παράμετρο σχήματος $\gamma = 0.15$

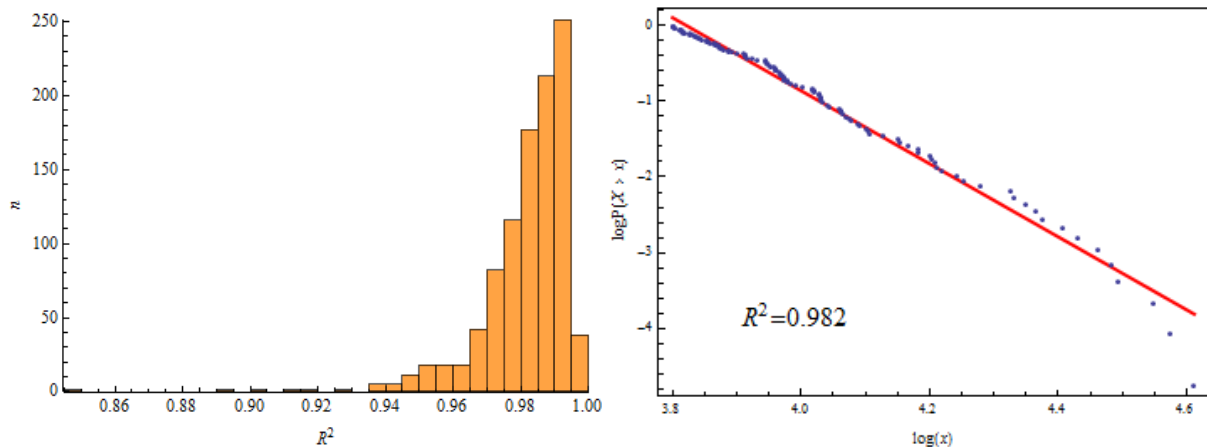
Όπως φαίνεται και από το Σχήμα 4.7, η κατανομή Pareto έχει υψηλές τιμές του συντελεστή R^2 . (Συγκεκριμένα μέση τιμή 0.982, τυπική απόκλιση 0.013, μέγιστη και ελάχιστη τιμή 0.998 και 0.871, τιμές παρεμφερείς με αυτές των εμπειρικών δεδομένων)



Σχήμα 4.8 Ιστόγραμμα τιμών συντελεστή R^2 για 1.000 τυχαία δείγματα κατανομής Pareto και γραφική παράσταση σε λογαριθμικούς άξονες για τη μέση τιμή του συντελεστή.

(β) Λογαριθμοκανονική κατανομή με παράμετρο κλίμακας $\beta = 2$ και παράμετρο σχήματος $\gamma = 0.8$

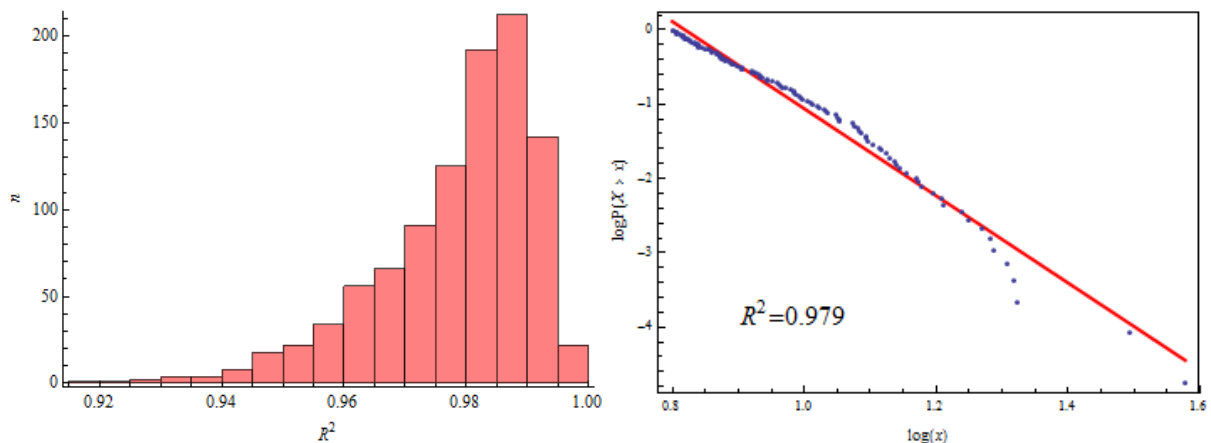
Και σ' αυτήν την περίπτωση ο συντελεστής R^2 παίρνει υψηλές τιμές, με μέση τιμή 0.982, τυπική απόκλιση 0.013 και μέγιστη και ελάχιστη τιμή 0.998 και 0.847.



Σχήμα 4.9 Ιστόγραμμα τιμών συντελεστή R^2 για 1.000 τυχαία δείγματα Λογαριθμοκανονικής κατανομής και γραφική παράσταση σε λογαριθμικούς άξονες για τη μέση τιμή του συντελεστή.

(γ) Εκθετική Κατανομή με παράμετρο $\beta = 2$.

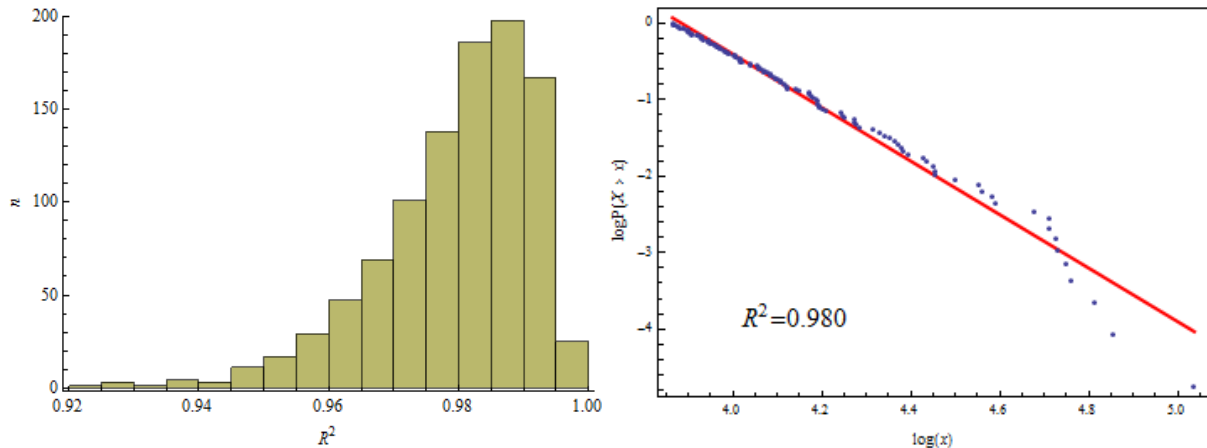
Ακόμα και στην περίπτωση της Εκθετικής κατανομής, όπου αναμένονταν μικρότερες τιμές του συντελεστή R^2 , και εδώ τα αποτελέσματα δε φαίνεται να διαφοροποιούνται σημαντικά (μέση τιμή 0.979, τυπική απόκλιση 0.013, μέγιστη και ελάχιστη τιμή 0.998 και 0.916)



Σχήμα 4.10 Ιστόγραμμα τιμών συντελεστή R^2 για 1.000 τυχαία δείγματα Εκθετικής κατανομής και γραφική παράσταση σε λογαριθμικούς άξονες για τη μέση τιμή του συντελεστή.

(δ) Κατανομή Weibull με παράμετρο κλίμακας $\beta = 6$ και παράμετρο σχήματος $\gamma = 0.7$

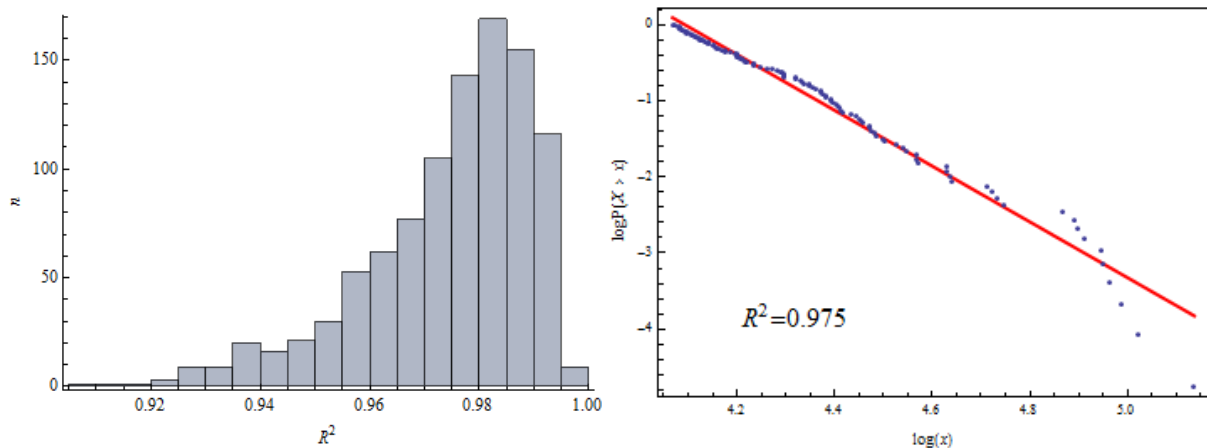
Παρόμοια συμπεριφορά και εδώ με μέση τιμή συντελεστή 0.980, τυπική απόκλιση 0.012, μέγιστη και ελάχιστη τιμή 0.998 και 0.922.



Σχήμα 4.11 Ιστόγραμμα τιμών συντελεστή R^2 για 1.000 τυχαία δείγματα κατανομής Weibull και γραφική παράσταση σε λογαριθμικούς άξονες για τη μέση τιμή του συντελεστή.

(ε) Γάμμα κατανομή με παράμετρο κλίμακας $\beta = 26$ και παράμετρο σχήματος $\gamma = 0.25$

Προέκυψε μέση τιμή συντελεστή R^2 ίση με 0.975, τυπική απόκλιση 0.015, μέγιστη και ελάχιστη τιμή 0.997 και 0.906.



Σχήμα 4.12 Ιστόγραμμα τιμών συντελεστή R^2 για 1.000 τυχαία δείγματα Γάμμα κατανομής και γραφική παράσταση σε λογαριθμικούς άξονες για τη μέση τιμή του συντελεστή.

Λαμβάνοντας υπόψη τα αποτελέσματα από την προσομοίωση Monte Carlo, προκύπτει το συμπέρασμα ότι η μέθοδος της γραφικής παράστασης σε λογαριθμικούς άξονες δεν είναι δυνατόν να συστηματοποιηθεί και να ληφθούν αξιόπιστα συμπεράσματα αποκλειστικά βάσει του συντελεστή R^2 . Η εφαρμογή της μεθόδου σε θεωρητικές κατανομές για τις οποίες δεν ανεμενόταν γραμμική συμπεριφορά του γραφήματος, έδειξε ότι ακόμα και σ' αυτές τις περιπτώσεις ο συντελεστής είχε υψηλές τιμές και δεν ήταν δυνατό να δημιουργηθεί ένα διάστημα μέσα στο οποίο να μπορούν να ενταχθούν οι ομαλά κυμαινόμενες κατανομές. Από οπτική εξέταση των αποτελεσμάτων όμως, είναι εμφανές ότι αρκετά μεγάλο ποσοστό των σταθμών αποδίδεται με ευθεία στο διπλό λογαριθμικό γράφημα και οι περιπτώσεις αυτές συνοδεύονται από συντελεστή $R^2 > 0.99$ και άνω. Τέτοιες τιμές συντελεστή καλύπτονται από το 30% των σταθμών.

Η μέθοδος εφαρμόστηκε και στη περίπτωση που θεωρούμε ως ουρά το 10% των παρατηρήσεων, για να ελεγχθεί αν το μέγεθος του δείγματος επηρεάζει την αξιοπιστία της μεθόδου. Τα αποτελέσματα όμως δεν είχαν σημαντική διαφορά και αυτό ενισχύει τον παραπάνω ισχυρισμό.

4.3 ΜΕΘΟΔΟΣ MEAN EXCESS FUNCTION (Συνάρτηση Μέσης Υπέρβασης)

4.3.1 Θεωρητικό υπόβαθρο

Η Συνάρτηση Μέσης Υπέρβασης (Σ.Μ.Υ.) είναι ένα χρήσιμο και συχνά χρησιμοποιούμενο εργαλείο για τη διάκριση ουρών. Μελετήθηκε αρχικά το 1960 από τους Benktander και Segerdahl (Benktander & Segerdahl, 1960). Έστω X τυχαία μεταβλητή με συνάρτηση κατανομής F και δεξί ακραίο σημείο x_f . Για ορισμένο $u < x_f$, η

$$F_u(x) = P(X - u \leq x | X > u), \quad x \geq 0 \quad (4.7)$$

είναι η συνάρτηση υπέρβασης της τυχαίας μεταβλητής X πάνω από το όριο u . Η συνάρτηση

$$e(u) = E(X - u | X > u) \quad (4.8)$$

καλείται mean excess function, ή συνάρτηση μέσης υπέρβασης της X .

Η μέθοδος βασίζεται στο γεγονός ότι η συνάρτηση $e(u)$ είναι σταθερή για κατανομές με εκθετική ουρά λόγω της αμνήμονης ιδιότητας της Εκθετικής κατανομής. Δηλαδή αν η X ακολουθεί Εκθετική κατανομή με παράμετρο β , τότε

$$e(u) = \beta \quad (4.9)$$

για όλα τα $u > 0$ ενώ για δυναμικού τύπου κατανομές με δείκτη ουράς α (με $\alpha > 2$ ή αντίστοιχα $\gamma < 1/2$) είναι

$$e(u) = \frac{u}{(a-2)} \quad (4.10)$$

Επομένως όταν σχεδιάζεται η εμπειρική τιμή του $e(u)$ συναρτήσει του u , εάν η γραφική παράσταση είναι γραμμική και η κλίση είναι ίση με μηδέν, αυτό συνιστά εκθετική κατανομή. Αν από την άλλη η γραφική παράσταση είναι γραμμική με κλίση μεγαλύτερη του μηδενός και η αποτέμνουσα είναι μηδέν, τότε έχουμε την περίπτωση υπο-εκθετικής κατανομής. Τέλος, στην περίπτωση που η κλίση είναι μικρότερη του μηδενός, η κατανομή είναι υπερεκθετική. Με τη μέθοδο αυτή λοιπόν, μπορεί να πραγματοποιηθεί διάκριση των κατανομών που ανήκουν στην τάξη D , με αυτές που ανήκουν στην τάξη E/D .

Έστω X τ.μ. μη φραγμένη στα δεξιά με συνάρτηση κατανομής F . Αν για κάθε $y \in \mathbb{R}$

$$\lim_{x \rightarrow \infty} \frac{\bar{F}(x-y)}{\bar{F}(x)} = e^{\gamma y} \quad (4.11)$$

για κάποιο $\gamma \in [0, \infty]$, τότε αποδεικνύεται ότι $\lim_{u \rightarrow \infty} e(u) = \gamma^{-1}$. Παρατηρούμε ότι η περίπτωση για $F \in \mathcal{S}$ (κατά Embrechts et.al) ή $F \in \mathcal{D}$ (κατά Ouarda et al.), δηλαδή η τάξη των υποεκθετικών κατανομών, ικανοποιείται όταν $\gamma = 0$ έτσι σ' αυτήν την περίπτωση των κατανομών με χονδρές ουρές, η συνάρτηση μέσης υπέρβασης $e(u)$ τείνει στο άπειρο καθώς το u τείνει στο άπειρο. Από την άλλη, υπερεκθετικές συναρτήσεις του τύπου $\bar{F}(x) \sim \exp(-x^a)$, $a > 1$ ικανοποιούν τη Σχέση 4.11 με $\gamma = \infty$ έτσι ώστε η συνάρτηση μέσης υπέρβασης να τείνει στο μηδέν (Embrechts et al., 1997).

Με τη μέθοδο είναι δυνατή η διάκριση μεταξύ μοντέλων με χοντρές και λεπτές ουρές. Χρειάζεται ιδιαίτερη προσοχή στην ερμηνεία αυτών των γραφημάτων. Εξαιτίας της

αραιώσης των δεδομένων που είναι διαθέσιμα για τον υπολογισμό του $e(u)$ για μεγάλες τιμές του u , τα γραφήματα που προκύπτουν είναι πολύ ευαίσθητα σε αλλαγές των δεδομένων προς το τέλος της σειράς. Γι' αυτό το λόγο, υπάρχουν και κάποιες εναλλακτικές εκδοχές όπως τα median excess plots (Embrechts et al., 1997)

4.3.2 Εφαρμογή μεθόδου σε βασικές θεωρητικές κατανομές

Χρησιμοποιώντας τον ορισμό της συνάρτησης μέσης υπέρβασης και με μερική ολοκλήρωση προκύπτει εύκολα ο ακόλουθος τύπος, που χρησιμοποιείται για να υπολογιστεί η μέση πιθανότητα υπέρβασης. Υποθέτουμε, για ευκολία αναπαράστασης, ότι η X είναι θετική τυχαία μεταβλητή με συνάρτηση κατανομής F και πεπερασμένη αναμονή. Μικρές αλλαγές επιτρέπουν όριο (x_0, ∞) για κάποιο $x_0 > 0$. Τότε

$$e(u) = \int_u^{x_F} (x-u) dF(x) / \bar{F}(u) = \frac{1}{\bar{F}(u)} \int_u^{x_F} \bar{F}(x) dx, \quad 0 < u < x_F \quad (4.12)$$

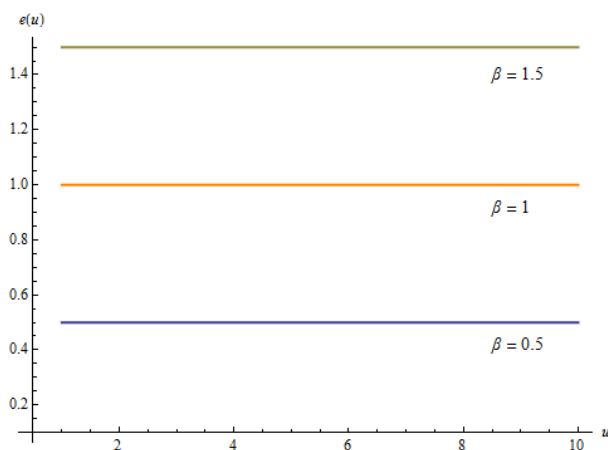
Όταν μάλιστα η F είναι συνεχής

$$\bar{F}(x) = \frac{e(0)}{e(x)} \exp \left\{ - \int_0^x \frac{1}{e(u)} du \right\}, \quad x > 0 \quad (4.13)$$

Συνεπάγεται αμέσως από τη Σχέση 4.13 ότι μια συνεχής συνάρτηση κατανομής ορίζεται μοναδικά από της συνάρτηση μέσης υπέρβασης αυτής.

Εφαρμόζοντας το θεωρητικό τύπο 4.12 προκύπτουν για τις βασικές κατανομές τα ακόλουθα γραφήματα που δίνουν τη μορφή της συνάρτησης μέσης υπέρβασης για κάθε περίπτωση.

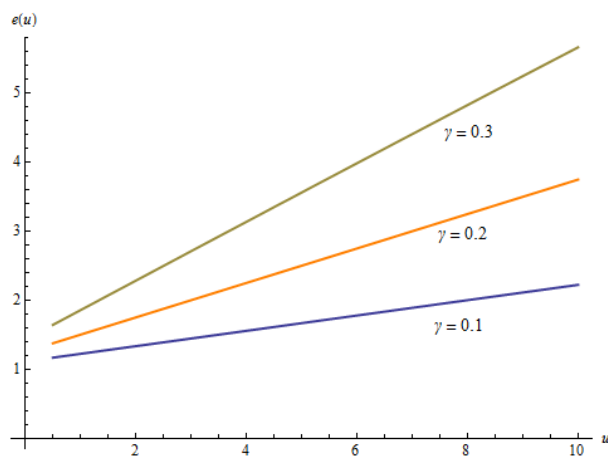
(α) Εκθετική Κατανομή



Σχήμα 4.13 Συνάρτηση Μέσης Υπέρβασης Εκθετικής κατανομής για διάφορες τιμές της παραμέτρου β .

Λόγω της αμνήμονης ιδιότητας της Εκθετικής κατανομής, για όλες τις παραμέτρους η Σ.Μ.Υ. είναι σταθερή για την κατανομή αυτή (Σχήμα 4.13).

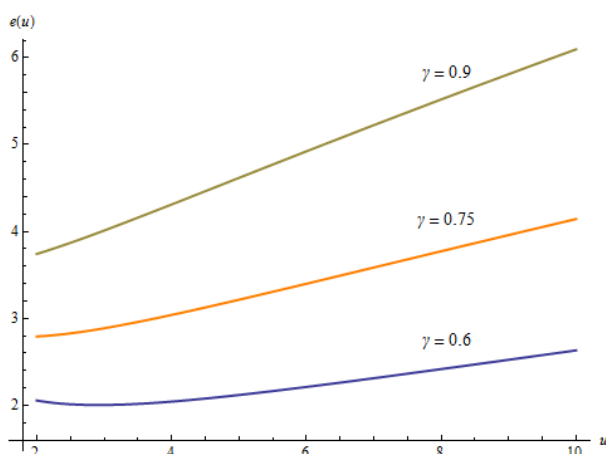
(β) Κατανομή Pareto τύπου II



Σχήμα 4.14 Συνάρτηση Μέσης Υπέρβασης κατανομής Pareto για διάφορες τιμές της παραμέτρου σχήματος γ .

Η κατανομή Pareto τύπου II, της οποίας οι ουρές χαρακτηρίζονται ως χονδρές, έχει αύξουσα Σ.Μ.Υ. (Σχήμα 4.14).

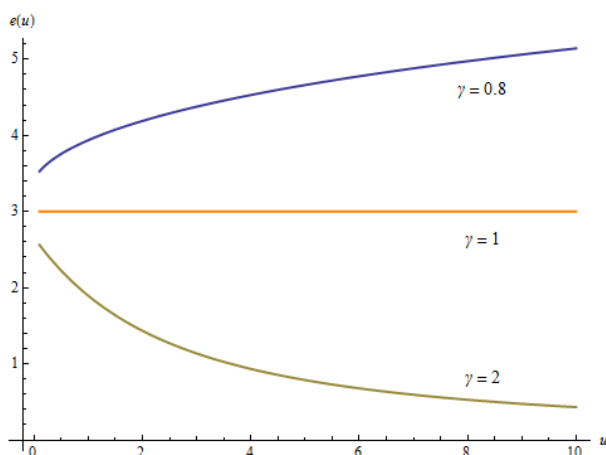
(γ) Λογαριθμοκανονική κατανομή



Σχήμα 4.15 Συνάρτηση Μέσης Υπέρβασης Λογαριθμοκανονικής κατανομής για διάφορες τιμές της παραμέτρου σχήματος γ .

Όπως η κατανομή Pareto, έτσι και η Λογαριθμοκανονική που ανήκει στην κατηγορία κατανομών με χονδρές ουρές έχει αύξουσα Σ.Μ.Υ. (Σχήμα 4.15)

(δ) Κατανομή Weibull

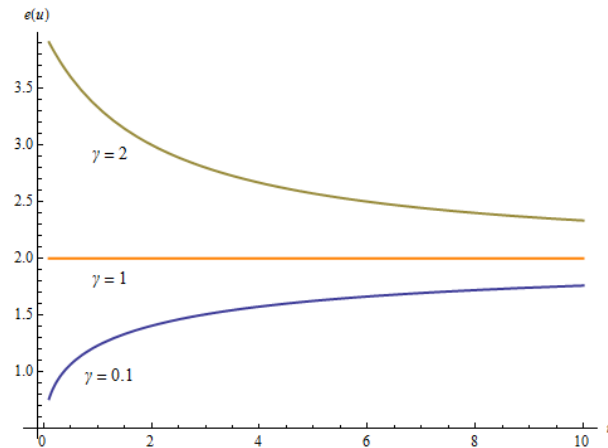


Σχήμα 4.16 Συνάρτηση Μέσης Υπέρβασης κατανομής Weibull για διάφορες τιμές της παραμέτρου σχήματος γ .

Για τιμές της παραμέτρου σχήματος γ μεγαλύτερες της μονάδας, η συνάρτηση είναι φθίνουσα, όπως φαίνεται στο Σχήμα 4.16, γεγονός που συνεπάγεται ουρές πιο λεπτές από αυτήν της Εκθετικής (υπερεκθετική κατανομή), ενώ αντίθετα για τιμές μικρότερες αυτής η

συνάρτηση είναι αύξουσα με χονδρές ουρές. Για παράμετρο σχήματος ίση με τη μονάδα η κατανομή εκφυλίζεται σε Εκθετική.

(ε) Κατανομή Γάμμα



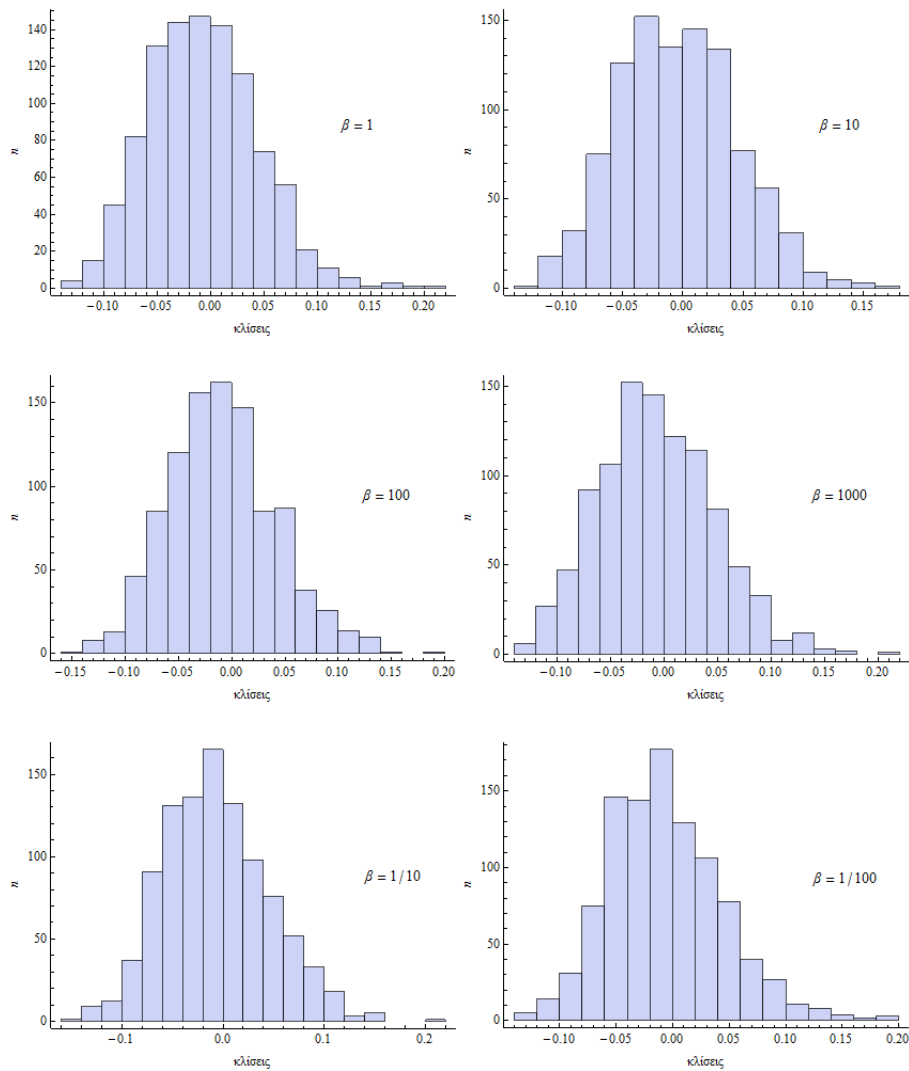
Σχήμα 4.17 Συνάρτηση Μέσης Υπέρβασης κατανομής Γάμμα για διάφορες τιμές της παραμέτρου σχήματος γ .

Η συνάρτηση Γάμμα, σαν ασυμπτωτικά Εκθετική, έχει μηδενική κλίση η οποία σταθεροποιείται με την αύξηση του u (Σχήμα 4.17).

4.3.3 Προσομοίωση Monte Carlo

- Παράμετρος Εκθετικής Κατανομής

Αρχικά πραγματοποιήθηκε έλεγχος για το βαθμό που επηρεάζονται τα διαστήματα των κλίσεων του γραφήματος Συνάρτησης Μέσης Υπέρβασης της Εκθετικής Κατανομής από την αλλαγή της παραμέτρου β . Για καθεμία από τις παραμέτρους $\beta = 1, 10, 100, 1000, 1/10, 1/100$ παρήχθησαν 1 000 τυχαία δείγματα Εκθετικής Κατανομής μήκους 10 000 τιμών στα οποία εφαρμόστηκε η μέθοδος. Σε αυτά θεωρήθηκε ως ουρά το μεγαλύτερο 10% των παρατηρήσεων. Προέκυψαν τα αποτελέσματα που φαίνονται στο Σχήμα 4.18.



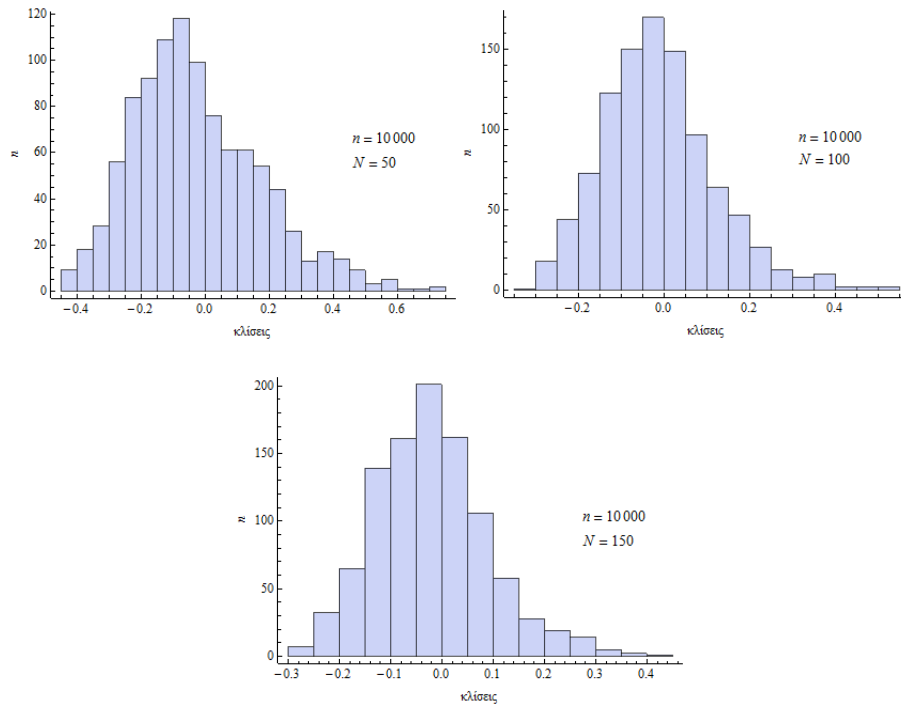
Σχήμα 4.18 Ιστογράμματα κλίσεων Σ.Μ.Υ. Εκθετικής κατανομής για διάφορες τιμές της παραμέτρου β .

Για όλες τις περιπτώσεις παραμέτρων, οι κλίσεις κυμαίνονται στο ίδιο διάστημα $[-0.1, 0.1]$ για διάστημα εμπιστοσύνης 95% με επικρατέστερη τιμή, όπως ήταν φυσικό, γύρω από το μηδέν. Επομένως, η επιλογή παραμέτρου Εκθετικής Κατανομής δεν επηρεάζει τα αποτελέσματα.

- Μήκος δείγματος και ουράς

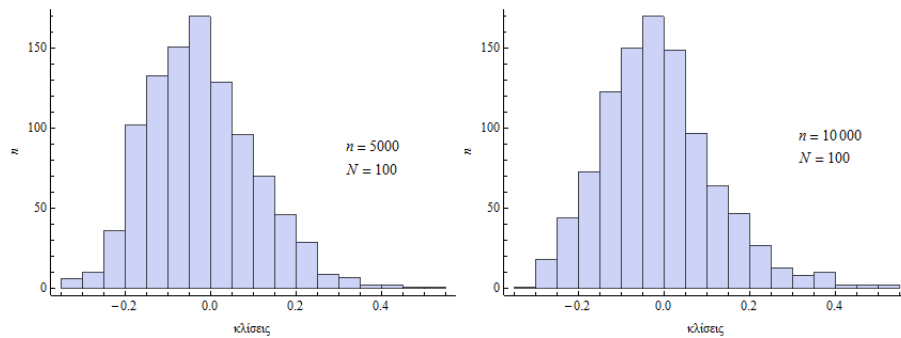
Ερευνήθηκε ο βαθμός στον οποίο επηρεάζονται οι κλίσεις από το μήκος του δείγματος n σε συνδυασμό με το πλήθος N των ακραίων τιμών που θεωρήθηκαν ως ουρά. Δοκιμάστηκαν

διάφοροι συνδυασμοί μηκών δείγματος και ουράς από τους οποίους επιβεβαιώθηκε η αμνήμονη ιδιότητα της Εκθετικής Κατανομής. Προέκυψε ως συμπέρασμα ότι για σταθερό μέγεθος ουράς, η κλίση παρέμενε στα ίδια επίπεδα ανεξάρτητα από το αρχικό μέγεθος του δείγματος. Στο Σχήμα 4.19 παρουσιάζεται η σύγκριση δείγματος σταθερού μεγέθους με μεταβαλλόμενο πλήθος τιμών ουράς.



Σχήμα 4.19 Ιστογράμματα κλίσεων Σ.Μ.Υ. Εκθετικής κατανομής για σταθερό δείγμα αλλά μεταβαλλόμενο μέγεθος ουράς.

Συγκρίνοντας τα διαγράμματα στο Σχήμα 4.19 προκύπτει ότι μεταβάλλοντας το πλήθος των τιμών των δεδομένων ουράς, το διάστημα των κλίσεων παρουσιάζει σημαντικές αλλαγές, και συγκεκριμένα με την αύξηση του αριθμού των τιμών ουράς, όπως είναι φυσικό, αυτό μειώνεται. Στο Σχήμα 4.20 συγκρίνονται δείγματα διαφορετικού μεγέθους αλλά ίσου μεγέθους ουράς



Σχήμα 4.20 Ιστογράμματα κλίσεων Σ.Μ.Υ. Εκθετικής κατανομής για σταθερό μήκος ουράς και μεταβαλλόμενο μέγεθος δείγματος.

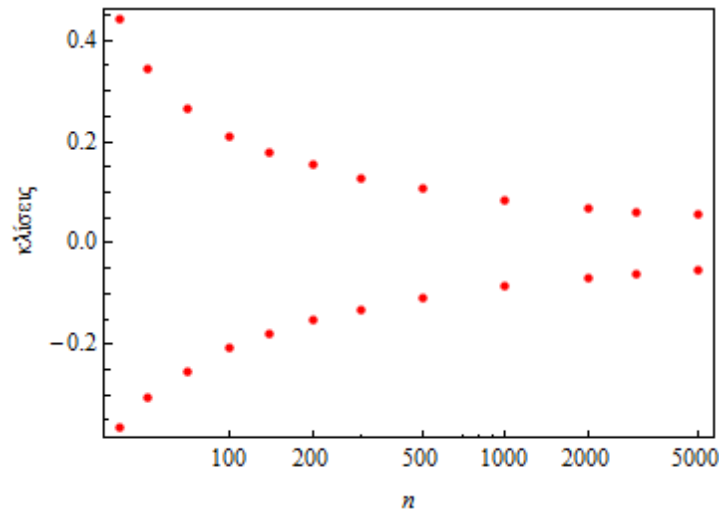
Είναι εμφανές (Σχήμα 4.20) ότι το διάστημα μέσα στο οποίο κυμαίνονται οι κλίσεις της Σ.Μ.Υ. είναι ανεξάρτητο από το μήκος του αρχικού δείγματος.

- Διαστήματα εμπιστοσύνης

Εφαρμόστηκαν δοκιμές Monte Carlo και πραγματοποιήθηκαν 10 000 προσομοιώσεις για να εξεταστεί η συμπεριφορά της Συνάρτησης Μέσης Υπέρβασης για την Εκθετική Κατανομή συναρτήσει του μήκους ουράς (οι δοκιμές έγιναν για παράμετρο $\beta = 1$, αφού όπως αποδείχθηκε η επιλογή παραμέτρου δεν έχει καμία επιρροή στα αποτελέσματα). Οι δοκιμές έγιναν για διάφορα μεγέθη δείγματος (από 40 έως 5 000).

Αφού έγινε εφαρμογή της Συνάρτησης Μέσης Υπέρβασης στα δεδομένα της θεωρητικής Εκθετικής Κατανομής, στα αποτελέσματα που προέκυψαν, εφαρμόστηκαν ευθείες και υπολογίστηκε η κλίση τους. Η διαδικασία αυτή πραγματοποιήθηκε για όλα τα μεγέθη δείγματος.

Επιπλέον, ορίστηκε διάστημα εμπιστοσύνης για τις κλίσεις που προέκυψαν ως αποτέλεσμα της εφαρμογής της Σ.Μ.Υ. στην Εκθετική κατανομή. Έτσι, για διάστημα εμπιστοσύνης 90%, το ακόλουθο γράφημα (Σχήμα 4.21) παρουσιάζει σε ημιλογαριθμικούς άξονες το διάστημα των κλίσεων για διάφορα μήκη δείγματος:



Σχήμα 4.21 Διάστημα εμπιστοσύνης 90% των κλίσεων της Σ.Μ.Υ. από την εφαρμογή σε Εκθετική κατανομή.

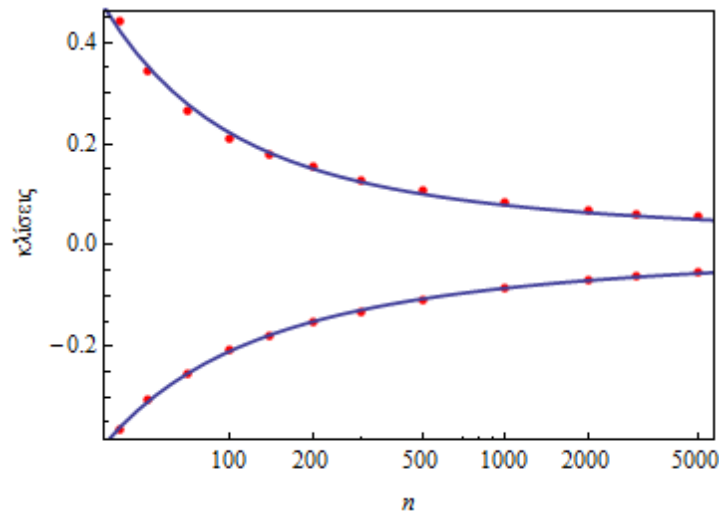
Επειδή το μήκος του δείγματος ποικίλλει, προσαρμόστηκε μια θεωρητική εξίσωση στα παραπάνω γραφήματα ώστε να εκφράζονται οι κλίσεις συναρτήσει του μήκους ουράς με αποτέλεσμα την ευχερέστερη επεξεργασία των πραγματικών δεδομένων. Χρησιμοποιώντας το μοντέλο με εξίσωση $y = b \log \left(1 + \frac{x}{d} \right)^{-c}$ προκύπτουν οι τιμές των συντελεστών που φαίνονται στον Πίνακα 4.1:

Πίνακας 4.1 Συντελεστές εξίσωσης ευθείας ορίων διαστήματος εμπιστοσύνης 90%

Δ.Ε.	όρια	b	c	d
90%	άνω όριο	0.4140	1.2531	23.866
	κάτω όριο	-0.9292	1.4957	7.1480

Τα δείγματα από την Εκθετική κατανομή που παρήχθησαν μέσω του προγράμματος είναι ανεξάρτητα μεταξύ τους. Αυτό όμως δεν ισχύει για τα πραγματικά δεδομένα βροχόπτωσης, όπου υπάρχει αυτοσυσχέτιση. Για το λόγο αυτό το τεστ κρίνεται αυστηρό στο αποτέλεσμά του, εξού και το μικρό διάστημα εμπιστοσύνης.

Στα παρακάτω γραφήματα φαίνεται η εφαρμογή του μοντέλου για δ.ε. 90% (Σχήμα 4.22).



Σχήμα 4.22 Μοντελοποίηση διαστήματος εμπιστοσύνης 90% των κλίσεων της Σ.Μ.Υ. από την εφαρμογή σε Εκθετική κατανομή.

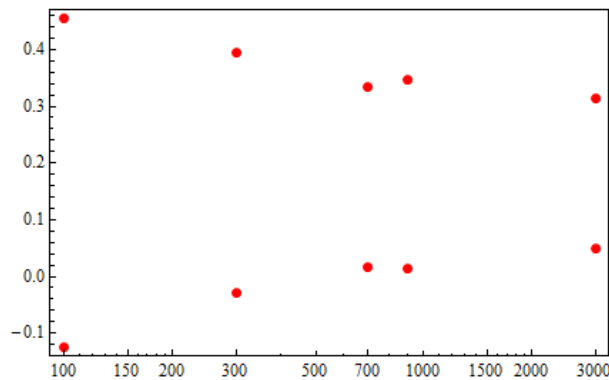
Είναι εφικτό πλέον να ελεγχθεί αν τα δεδομένα ακολουθούν Εκθετική Κατανομή, εφαρμόζοντας ευθεία στο γράφημα της Συνάρτησης Μέσης Υπέρβασής τους και εξετάζοντας αν η κλίση της ευθείας αυτής είναι μέσα στα όρια των διαστημάτων. Εάν η κλίση βρίσκεται πάνω από τα όρια, η ουρά είναι υποεκθετική, και αν βρίσκεται κάτω από αυτά υπερεκθετική. Η διαδικασία αυτή επιτυγχάνεται μέσω της κατασκευής ενός «τεστ εκθετικότητας» (βλ. Παράρτημα Β)

- Δοκιμές Monte Carlo για θεωρητικές κατανομές

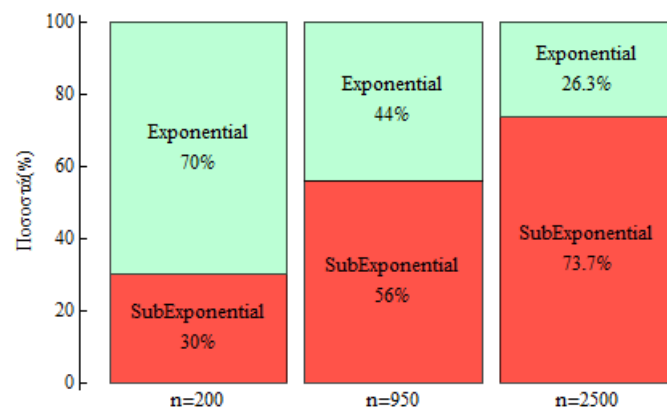
Παρουσιάζονται τα αποτελέσματα της προσομοίωσης Monte Carlo για τη μέθοδο της Συνάρτησης Μέσης Υπέρβασης. Έχουν παραχθεί 1.000 τυχαία δείγματα για κάποιες τιμές παραμέτρων των εξεταζόμενων κατανομών. Στα παραγόμενα αυτά δείγματα, εφαρμόστηκε η μέθοδος και το τεστ εκθετικότητας και για κάθε κατανομή προέκυψαν τα παρακάτω αποτελέσματα.. Στα παρακάτω διαγράμματα το n είναι το μέγεθος δείγματος ουράς του οποίου οι τιμές επιλέχθηκαν τέτοιες ώστε να είναι κοντά στην ελάχιστη, τη μέγιστη και τη μέση τιμή.

(α) Κατανομή Pareto τύπου II

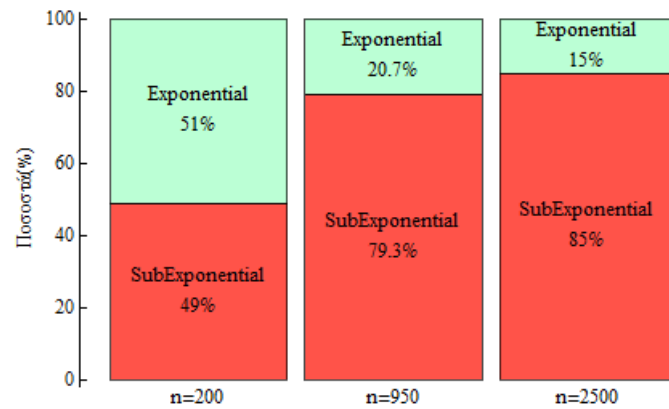
Για μια μέση παράμετρο σχήματος Γενικευμένης Κατανομής Pareto $\gamma = 0.15$ εκτιμήθηκαν για διάστημα εμπιστοσύνης 90% τα όρια των κλίσεων (για 1 000 τυχαία δείγματα ανά μήκος δείγματος ουράς) που φαίνονται στο παρακάτω διάγραμμα και αναφέρονται σε μήκη δείγματος που αντιστοιχούν στην ελάχιστη και μέγιστη τιμή, στο μέσο και τα ποσοστημόρια 0.25 και 0.75 των εμπειρικών δειγμάτων. Είναι φανερό ότι οι κλίσεις είναι καθαρά θετικές για μεγάλα μήκη δείγματος και συγκεκριμένα ουράς (1 000 και άνω)



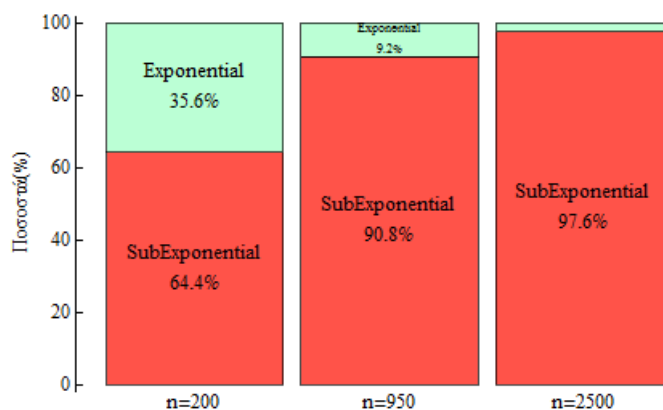
Σχήμα 4.22 Κλίσεις της Σ.Μ.Υ. για την κατανομή Pareto για διάστημα εμπιστοσύνης 90% και $\gamma = 0.15$.



Σχήμα 4.23 Αποτελέσματα τεστ Εκθετικότητας συναρτήσει του μεγέθους ουράς σε τυχαία δείγματα κατανομής Pareto για παράμετρο σχήματος $\gamma = 0.1$.



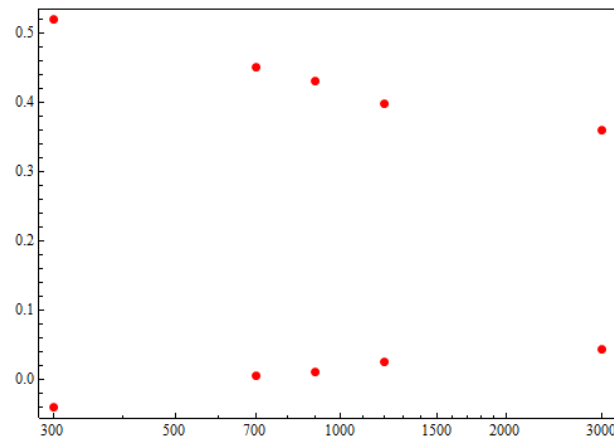
Σχήμα 4.24 Αποτελέσματα τεστ Εκθετικότητας συναρτήσε του μεγέθους ουράς σε τυχαία δείγματα κατανομής Pareto για παράμετρο σχήματος $\gamma = 0.15$.



Σχήμα 4.25 Αποτελέσματα τεστ Εκθετικότητας συναρτήσε του μεγέθους ουράς σε τυχαία δείγματα κατανομής Pareto για παράμετρο σχήματος $\gamma = 0.2$.

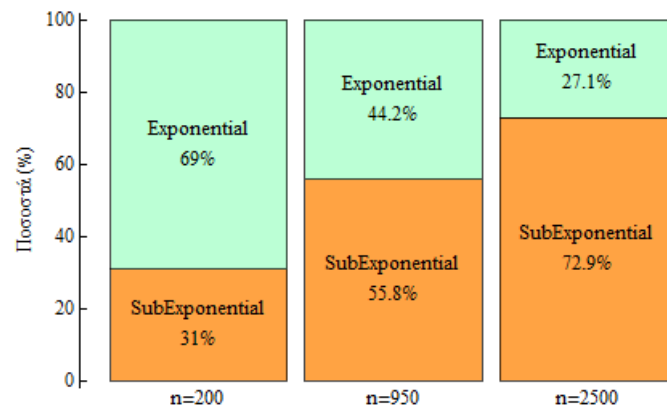
Το διάγραμμα διαστήματος εμπιστοσύνης (Σχήμα 4.22) δείχνει κλίσεις μεγαλύτερες του μηδενός. Για μικρή παράμετρο σχήματος κατανομής Pareto το τεστ είναι αποτελεσματικό μόνο για μεγάλο μήκος δείματος. Έτσι κι αλλιώς όμως το 80% των σταθμών που εξετάστηκε είχε μήκος δείματος μεγαλύτερο των 6 000, άρα είχε μήκος ουράς μεγαλύτερο από 600, επομένως τα αποτελέσματα κρίνονται αρκούντως ικανοποιητικά. Όσο βέβαια η παράμετρος αυξάνεται μπορούμε να θεωρήσουμε ότι έχουμε ικανοποιητικά αποτελέσματα ακόμα και για μικρό δείγμα (Σχήματα 4.23 έως 4.25).

(β) Λογαριθμοκανονική Κατανομή

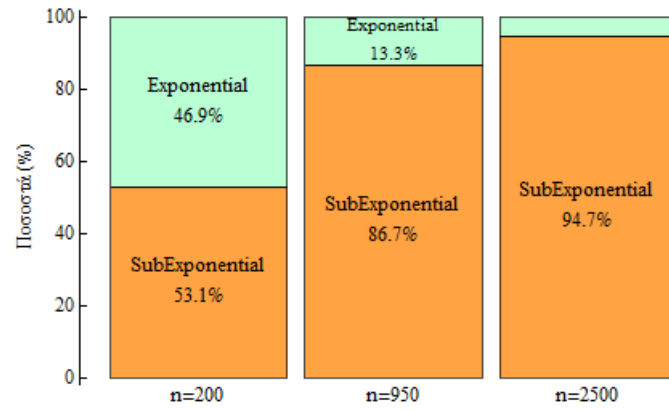


Σχήμα 4.26 Διάστημα εμπιστοσύνης 90% των κλίσεων της Σ.Μ.Υ. για τη Λογ/νική κατανομή για $\gamma = 0.8$

Σύμφωνα με το Σχήμα 4.26 και για τη Λογαριθμοκανονική κατανομή οι κλίσεις της Σ.Μ.Υ. παίρνουν τιμές μεγαλύτερες του μηδενός.



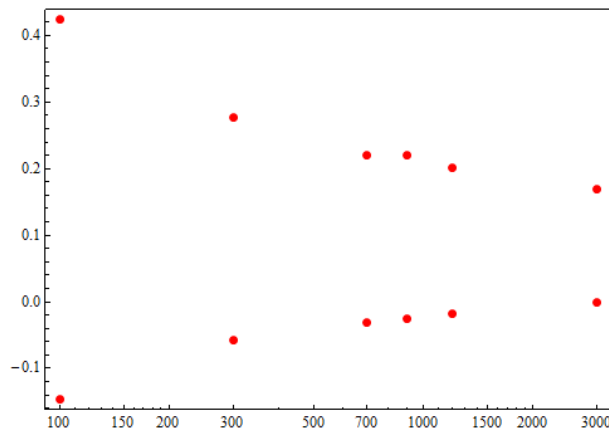
Σχήμα 4.27 Αποτελέσματα τεστ Εκθετικότητας συναρτήσεως του μεγέθους ουράς σε τυχαία δείγματα Λογαριθμοκανονικής κατανομής για παράμετρο σχήματος $\gamma = 0.6$.



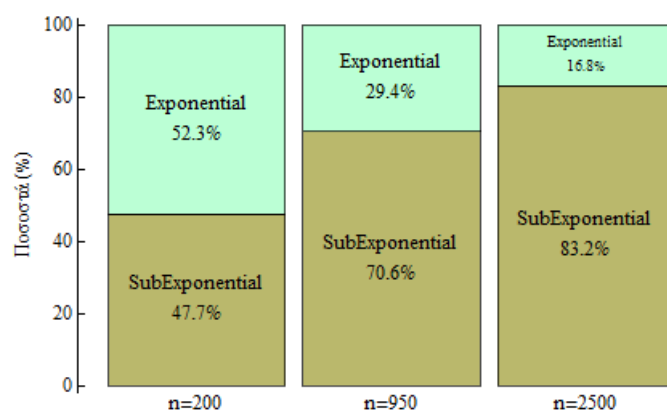
Σχήμα 4.28 Αποτελέσματα τεστ Εκθετικότητας συναρτήσεως του μεγέθους ουράς σε τυχαία δείγματα Λογαριθμοκανονικής κατανομής για παράμετρο σχήματος $\gamma = 0.8$.

Και για την περίπτωση της Λογαριθμοκανονικής κατανομής, για μικρή παράμετρο σχήματος χρειάζεται μεγάλο μήκος δείγματος για να θεωρηθεί αξιόπιστη η μέθοδος, ενώ όσο αυτή αυξάνεται τα αποτελέσματα είναι καλύτερα.

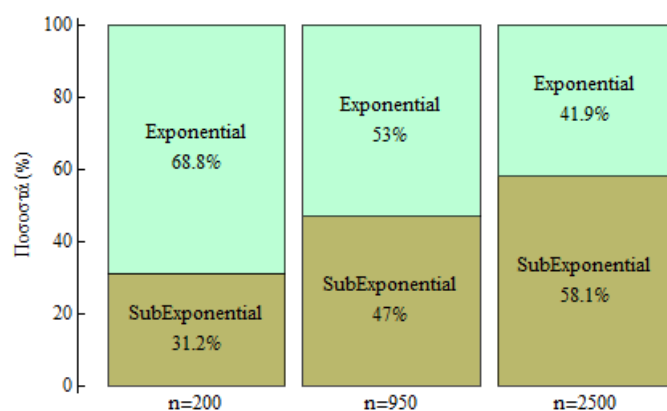
(γ) Κατανομή Weibull



Σχήμα 4.29 Διάστημα εμπιστοσύνης 90% των κλίσεων της Σ.Μ.Υ. για την κατανομή Weibull με $\gamma = 0.7$



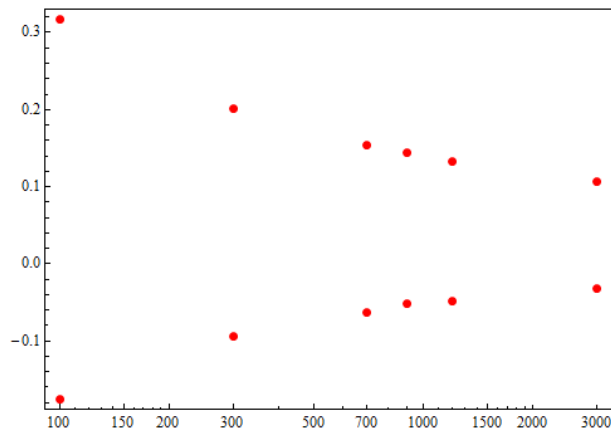
Σχήμα 4.30 Αποτελέσματα τεστ Εκθετικότητας συναρτήσεως του μεγέθους ουράς σε τυχαία δείγματα κατανομής Weibull για παράμετρο σχήματος $\gamma = 0.6$.



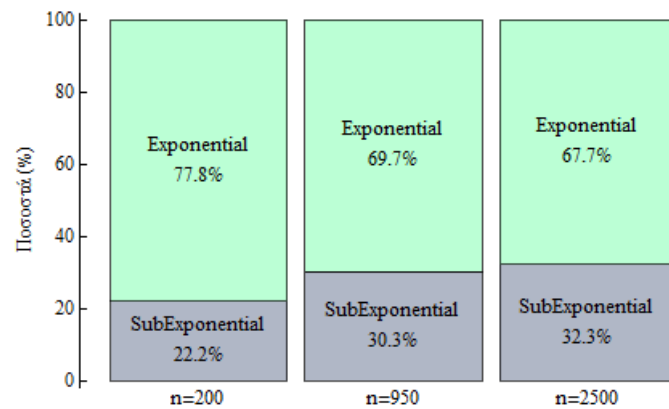
Σχήμα 4.31 Αποτελέσματα τεστ Εκθετικότητας συναρτήσεως του μεγέθους ουράς σε τυχαία δείγματα μεταβαλλόμενου μεγέθους κατανομής Weibull για παράμετρο σχήματος $\gamma = 0.8$.

Για την κατανομή Weibull η αύξηση της παραμέτρου σχήματος συνεπάγεται «αύξηση» εκθετικότητας (Σχήμα 4.30). Για παράμετρο ίση με 0.8, που είναι μια ενδιάμεση κατάσταση, τα αποτελέσματα δεν είναι ξεκάθαρα όπως ήταν αναμενόμενο (Σχήμα 4.31).

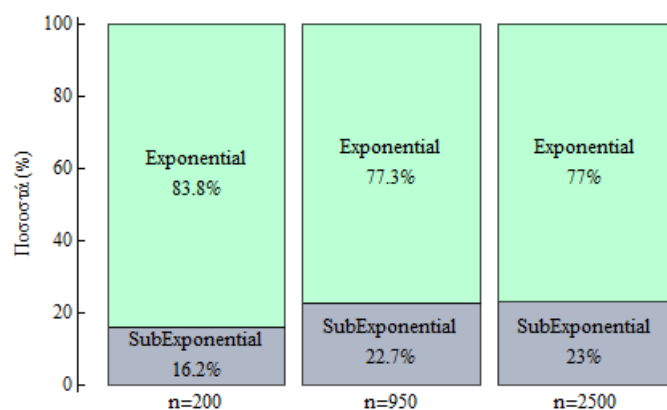
(δ) Κατανομή Γάμμα



Σχήμα 4.32 Διάστημα εμπιστοσύνης 90% των κλίσεων της Σ.Μ.Υ. για την κατανομή Weibull με $\gamma = 0.7$



Σχήμα 4.33 Αποτελέσματα τεστ Εκθετικότητας συναρτήσει του μεγέθους ουράς σε τυχαία δείγματα κατανομής Γάμμα για παράμετρο σχήματος $\gamma = 0.2$.



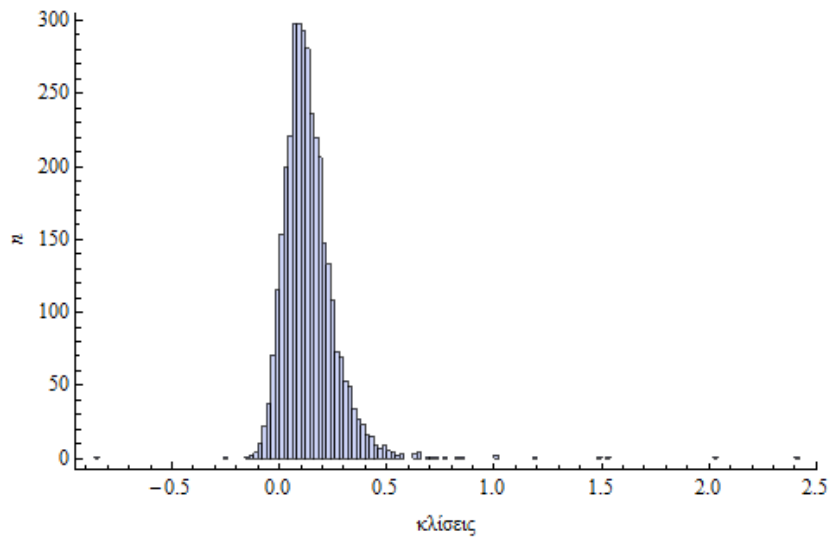
Σχήμα 4.34 Αποτελέσματα τεστ Εκθετικότητας συναρτήσεως του μεγέθους ουράς σε τυχαία δείγματα κατανομής Γάμμα για παράμετρο σχήματος $\gamma = 0.3$.

Η Γάμμα, όπως ήταν αναμενόμενο, έχει υψηλά ποσοστά «εκθετικότητας» (Σχήματα 4.33 και 4.34). Επιπλέον τα Σχήματα 4.29 και 4.32 δείχνουν ότι οι κατανομές Weibull και Γάμμα έχουν μέγιστες τιμές κλίσεων μικρότερες των Pareto και Λογαριθμοκανονικής.

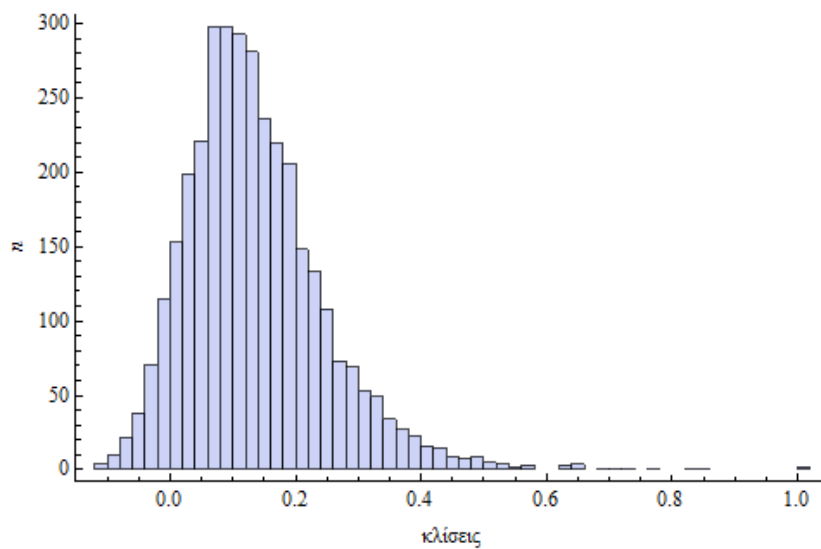
4.3.4 Εφαρμογή Μεθόδου στα δεδομένα

Για την εφαρμογή της μεθόδου, βρέθηκε η μέση τιμή της δεσμευμένης πιθανότητας (σχέση 4.8) για τα δεδομένα της «ουράς» κάθε σταθμού και ως μέγιστο κατώφλι u επιλεγόταν κάθε φορά η 20^{th} μεγαλύτερη τιμή του δείγματος, ώστε να υπάρχουν τουλάχιστον 20 τιμές που να υπερβαίνουν το u . Ως ουρά θεωρήθηκε το μεγαλύτερο 10% των παρατηρήσεων. Η επιλογή του μεγάλου μήκους ουράς έγινε γιατί όπως αναφέρθηκε η μέθοδος αποδίδει καλύτερα σε μεγάλα δείγματα (Embrechts et al., 1997). Έπειτα, έγινε προσαρμογή ευθείας στο προκύπτον γράφημα με σκοπό την εύρεση της κλίσης αυτού.

Αφού εφαρμόστηκε ο κατάλληλος αλγόριθμος στα δεδομένα (βλ. Παράρτημα Β), βρέθηκε ότι αυτά έχουν κλίσεις γραφήματος Σ.Μ.Υ. με μέση τιμή 0.138, διάμεσο 0.121 και τυπική απόκλιση 0.13. Ακολουθεί ιστόγραμμα των δεδομένων με τις κλίσεις των διαγραμμάτων Σ.Μ.Υ. των σταθμών (Σχήματα 4.35 και 4.36).



Σχήμα 4.35 Ιστόγραμμα κλίσεων Σ.Μ.Υ. συνόλου σταθμών.



Σχήμα 4.36 Ιστόγραμμα κλίσεων Σ.Μ.Υ. σταθμών (έχουν αφαιρεθεί οι πέντε μεγαλύτερες και μικρότερες τιμές για μεγαλύτερη ευκρίνεια) .

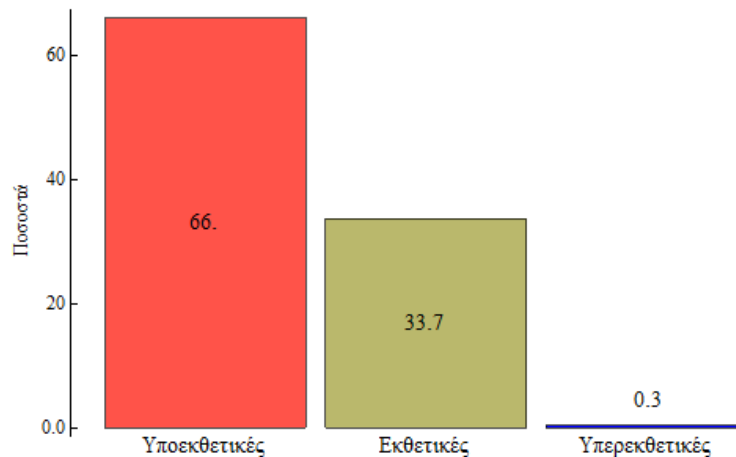
Κάποια στατιστικά χαρακτηριστικά των κλίσεων φαίνονται στον Πίνακα 4.2

Πίνακας 4.2 Στατιστικά χαρακτηριστικά κλίσεων Σ.Μ.Υ. σταθμών

Min	Q ₁	Διάμεσος	Q ₃	Max
-0.852	0.062	0.121	0.194	2.402

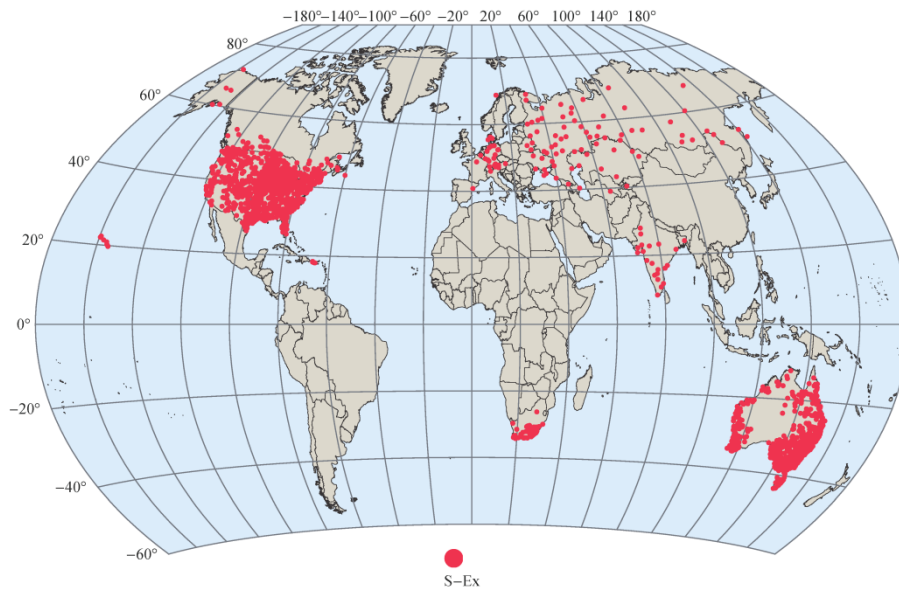
- Αποτελέσματα από την εφαρμογή των ελέγχων

Από την εφαρμογή του τεστ για διάστημα εμπιστοσύνης 90% προέκυψε ότι από τους 3 477 σταθμούς, οι 1 170 (δηλαδή ένα ποσοστό 33.5%) είχαν κλίσεις μέσα στα όρια στα οποία ορίσαμε την Εκθετική Κατανομή, επομένως η υπόθεση να έχουν εκθετική ουρά δεν μπορεί να απορριφθεί. Υπήρχαν επίσης και 10 σταθμοί με υπερεκθετικές ουρές (0.29%)

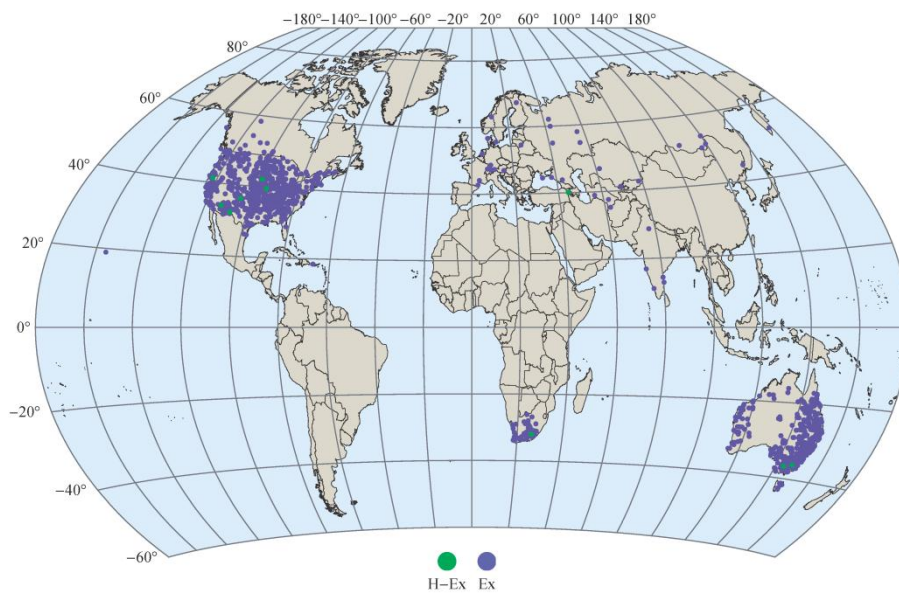


Σχήμα 4.37 Ιστόγραμμα ποσοστών εκθετικών, υποεκθετικών και υπερεκθετικών ουρών βάσει της Σ.Μ.Υ.

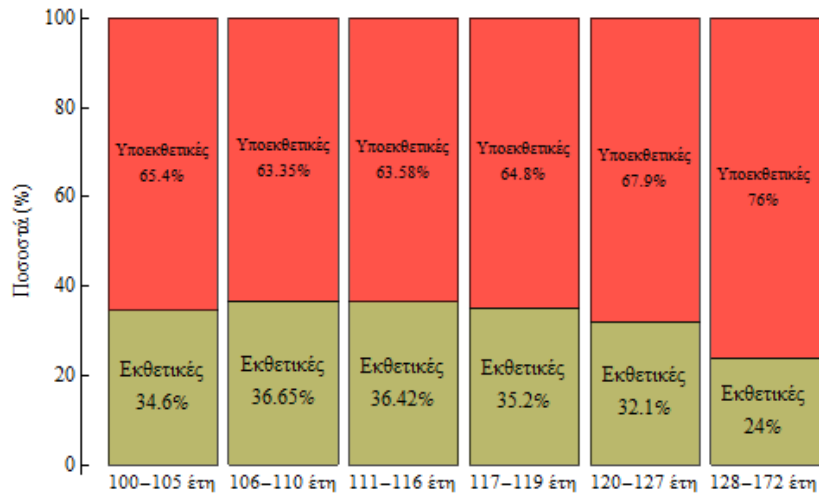
Ο παραπάνω έλεγχος (Σχήμα 4.37) αποδεικνύει ότι οι βροχοπτώσεις σε παγκόσμιο επίπεδο ακολουθούν (στην πλειοψηφία των σταθμών) υπο-εκθετικές κατανομές, έχουν δηλαδή χοντρές ουρές. Στα Σχήματα 4.38 και 4.39 φαίνεται η γεωγραφική κατανομή των υποεκθετικών και ξεχωριστά των εκθετικών και υπερεκθετικών κατανομών. Στο Σχήμα 4.40 φαίνεται η αναλογία εκθετικών – υποεκθετικών κατηγοριοποιημένα με βάση το μήκος των ετών των σταθμών. Η κατηγοριοποίηση των ετών έγινε με τον τρόπο που φαίνεται στο Σχήμα 4.40 με την προϋπόθεση σε κάθε κατηγορία να υπάρχουν 500-600 σταθμοί με πλήθος ετών που να ανήκει στην αντίστοιχη κατηγορία, με εξαίρεση την τελευταία (128-172 έτη με 380 σταθμούς). Για πλήθος ετών 111 και άνω το ποσοστό των εκθετικών ουρών μειώνεται με την αύξηση των ετών. Μάλιστα η διαφορά του ποσοστού της κατηγορίας των 111-116 έτη σε σχέση με το μέγιστο πλήθος ξεπερνά το 10%.



Σχήμα 4.38 Γεωγραφική κατανομή υποεκθετικών ουρών βάσει της μεθόδου Σ.Μ.Υ.



Σχήμα 4.39 Γεωγραφική κατανομή εκθετικών και υπερεκθετικών ουρών βάσει της μεθόδου Σ.Μ.Υ.



Σχήμα 4.40 Ποσοστά εκθετικών – υποεκθετικών ουρών συναρτήσει των ετών μέτρησης των σταθμών

4.4 ΜΕΘΟΔΟΣ ΕΛΑΧΙΣΤΟΠΟΙΗΣΗΣ ΣΦΑΛΜΑΤΟΣ

Με τη μέθοδο αυτή, προσαρμόστηκε και συγκρίθηκε η συμπεριφορά διαφόρων ουρών θεωρητικών κατανομών σε σχέση με τις εμπειρικές κατανομές που εκτιμήθηκαν από τα ημερήσια δεδομένα των σταθμών. Οι θεωρητικές ουρές προσαρμόζονται στις εμπειρικές μέσω της ελαχιστοποίησης μιας νόρμας μέσου τετραγωνικού σφάλματος.

4.4.1 Σύγκριση Τύπων ελαχιστοποίησης σφάλματος

Δοκιμάζουμε τέσσερα μοντέλα, των οποίων η απόδοση συγκρίνεται με σκοπό την επιλογή του καταλληλότερου. Οι τέσσερις νόρμες έχουν ως εξής:

$$\sum_{i=n-N+1}^n [x'_i - x_i]^2 \quad (4.14)$$

$$\sum_{i=n-N+1}^n \left[\frac{x'_i}{x_i} - 1 \right]^2 \quad (4.15)$$

$$\sum_{i=n-N+1}^n [\bar{F}(x'_i) - \bar{F}(x_i)]^2 \quad (4.16)$$

$$\sum_{i=n-N+1}^n \left[\frac{\bar{F}(x_i)}{\bar{F}(x'_i)} - 1 \right]^2 \quad (4.17)$$

Ο τύπος 4.14 είναι ο συνήθως χρησιμοποιούμενος τύπος μέσου τετραγωνικού σφάλματος με $x = Q(u)$, Q συνάρτηση για το ποσοστημόριο της κατανομής που μελετάται, και $u = \frac{i}{n+1}$.

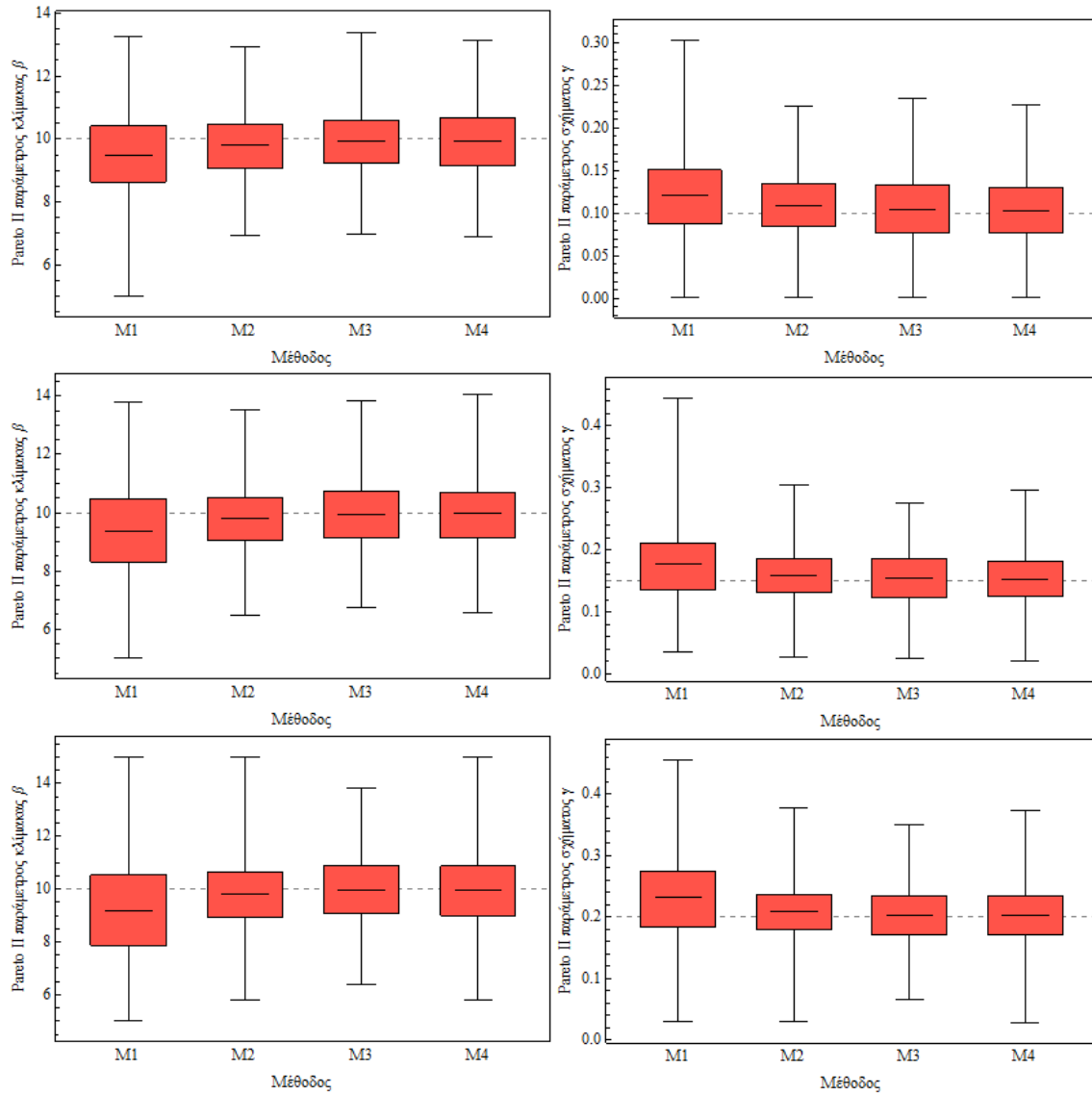
Με x'_i συμβολίζονται οι εμπειρικές τιμές. Για τις νόρμες (4.16) και (4.17) το $\bar{F}(x'_i)$ είναι η εμπειρική συνάρτηση υπέρβασης, που σύμφωνα με τη σχεδιαστική θέση κατά Weibull ισούται με $u = \frac{i}{n+1}$.

Για κάθε θεωρητική κατανομή δοκιμάζεται η αποτελεσματικότητα κάθε νόρμας. Η διαδικασία είναι κοινή. Επιλέγεται μέγεθος τιμών $N = 116$ για να οριστεί η ουρά, όσο είναι δηλαδή ο μέσος όρος των ετών των καταγραφών (βλ. Ορισμός ανώτατου μέρους κατανομής). Ορίζεται επίσης $m = 9\ 500$ ο μέσος όρος των θετικών τιμών των δεδομένων των σταθμών.

Για κάθε βασική θεωρητική κατανομή εφαρμόζονται οι νόρμες για διάφορες παραμέτρους σχήματος γ . Οι νόρμες αυτές δίνουν το σφάλμα μεταξύ εμπειρικών δεδομένων και κατανομής που προσαρμόζεται. Στην προκειμένη περίπτωση τα εμπειρικά δεδομένα είναι τυχαία παραγόμενα δείγματα από τις υπό εξέταση ουρές κατανομών για κάποιες τιμές παραμέτρων και η νόρμα ελαχιστοποιείται για να προκύψουν παράμετροι ίσες με αυτές που θεωρήθηκαν αρχικά. Ελαχιστοποιώντας κάθε τύπο έχουμε ξεχωριστά τα παρακάτω διασαφηνιστικά θηκογράμματα.

(α) Κατανομή Pareto με παράμετρο κλίμακας $\beta = 10$

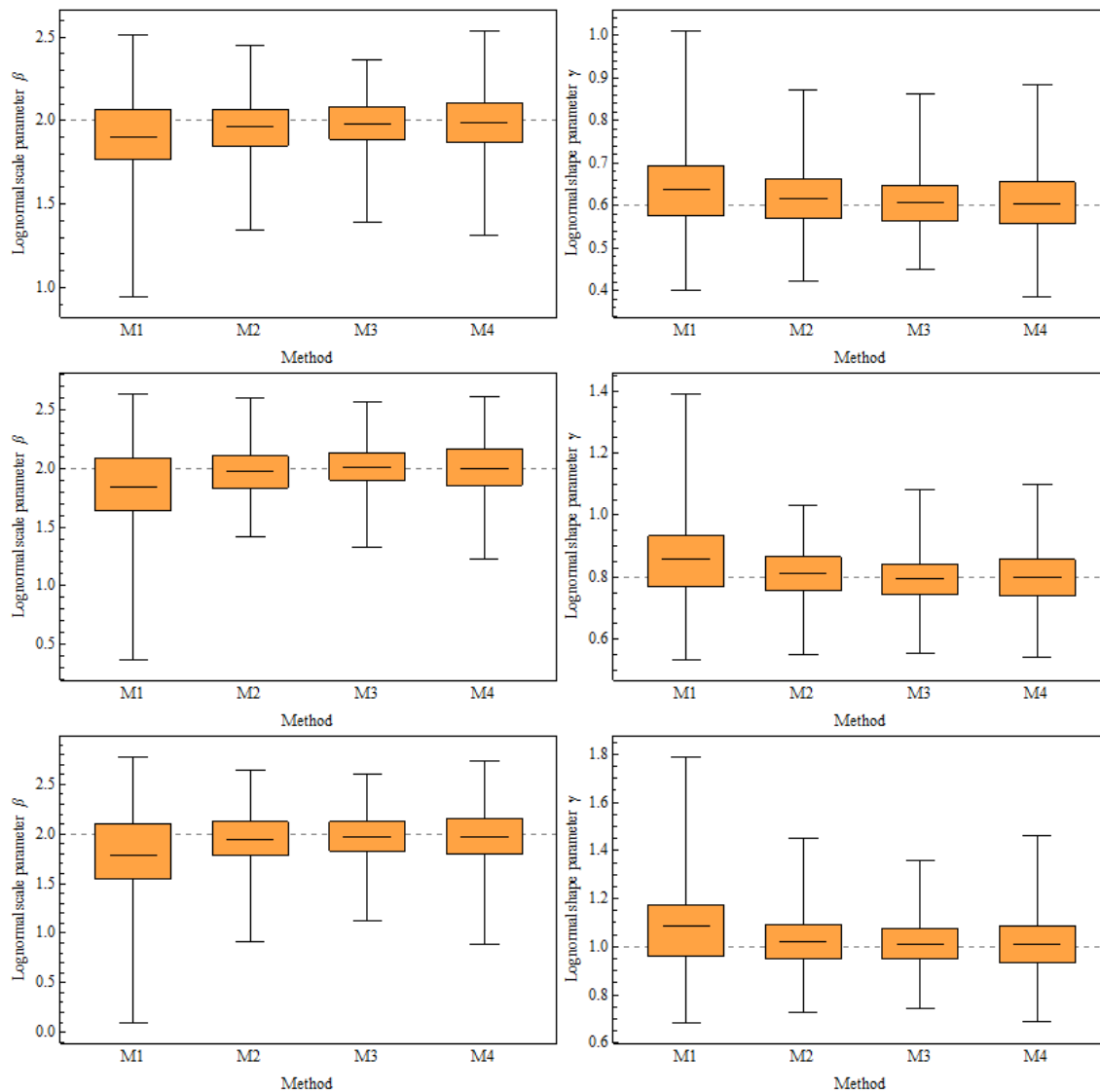
Οι τέσσερις νόρμες δοκιμάστηκαν για παράμετρο σχήματος ίση με 0.1, 0.15 και 0.2. Τα αποτελέσματα φαίνονται στο Σχήμα 4.41



Σχήμα 4.41 Δοκιμές μοντέλων στην κατανομή Pareto για β ίση με 10 και γ 0.1, 0.15 και 0.2 αντίστοιχα. Η νόρμα 4.17 έχει τα καλύτερα αποτελέσματα

(β) Λογαριθμοκανονική Κατανομή με παράμετρο θέσης $\beta = 2$

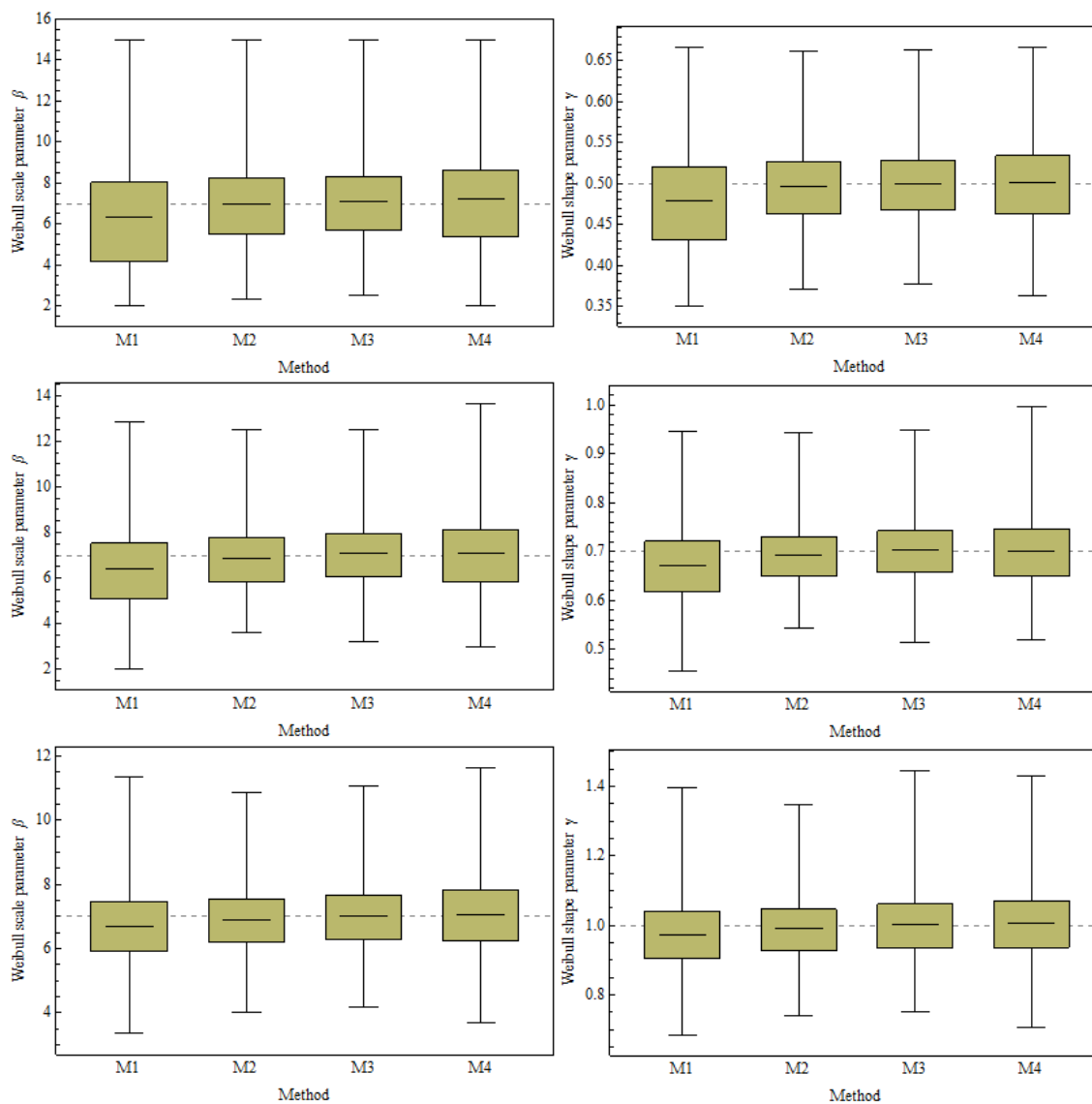
Οι τέσσερις νόρμες δοκιμάστηκαν για παράμετρο σχήματος ίση με 0.6, 0.8 και 1. Τα αποτελέσματα φαίνονται στο Σχήμα 4.42



Σχήμα 4.42 Δοκιμές μοντέλων στη Λογαριθμοκανονική κατανομή για β ίση με 2 και γ 0.6, 0.8 και 1 αντίστοιχα. Η νόρμα 4.17 έχει τα καλύτερα αποτελέσματα.

(γ) Κατανομή Weibull με παράμετρο κλίμακας $\beta = 7$

Οι τέσσερις νόρμες δοκιμάστηκαν για παράμετρο σχήματος ίση με 0.5, 0.7 και 1. Τα αποτελέσματα φαίνονται στο Σχήμα 4.43

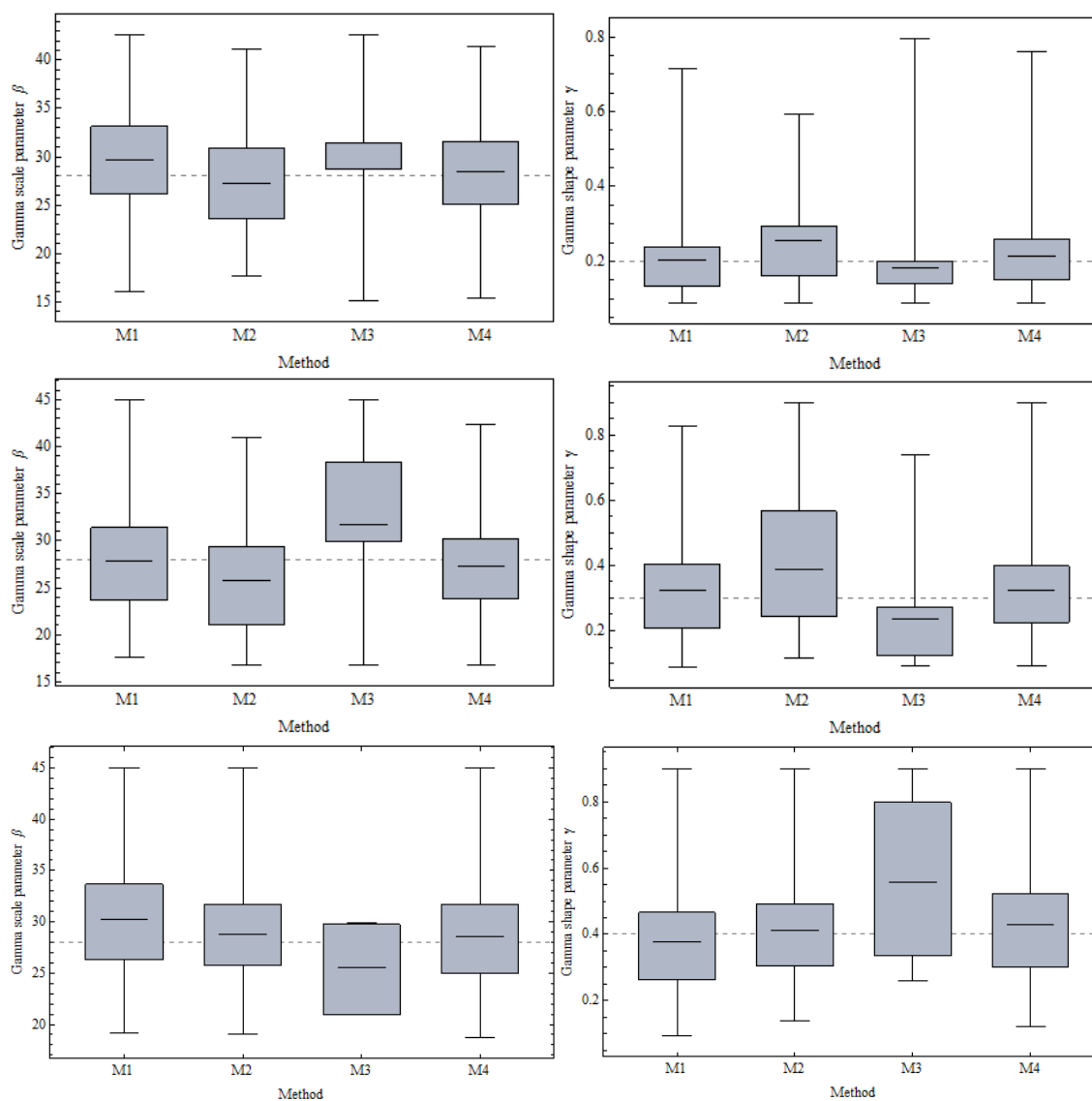


Σχήμα 4.43 Δοκιμές μοντέλων στη Weibull κατανομή για β ίση με 2 και γ 0.5, 0.7 και 1 αντίστοιχα. Η νόρμα 4.17 έχει τα καλύτερα αποτελέσματα.

Αξίζει να σημειωθεί ότι με τη μείωση του συντελεστή γ της Weibull, μειώνεται και η απόδοση της νόρμας 4.14.

(δ) Κατανομή Γάμμα με παράμετρο θέσης $\beta = 28$

Οι τέσσερις νόρμες δοκιμάστηκαν για παράμετρο σχήματος ίση με 0.2, 0.3 και 0.4. Τα αποτελέσματα φαίνονται στο Σχήμα 4.44



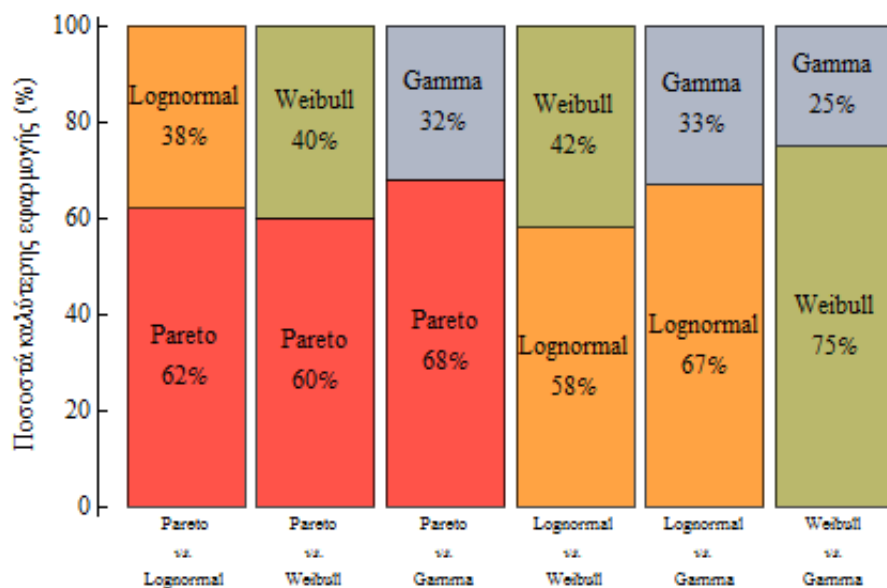
Σχήμα 4.44 Δοκιμές μοντέλων στη Γάμμα κατανομή για β ίση με 2 και γ 0.2, 0.3 και 0.4 αντίστοιχα. Η νόρμα 4.17 φαίνεται να έχει τα καλύτερα αποτελέσματα.

Σε όλες τις περιπτώσεις είναι προφανές ότι καλύτερο αποτέλεσμα δίνει η νόρμα 4.17, έχοντας την καλύτερη προσέγγιση και για τις δύο παραμέτρους, άρα θα είναι και αυτή που θα χρησιμοποιηθεί. Αξίζει να σημειωθεί ότι αν και στις κατανομές Pareto, Λογαριθμοκανονική

και Weibull η δεύτερη καλύτερη νόρμα ήταν η 4.16, στην περίπτωση της κατανομής Γάμμα δίνει τη χειρότερη προσέγγιση παραμέτρων.

4.4.2 Εφαρμογή Μεθόδου στα δεδομένα

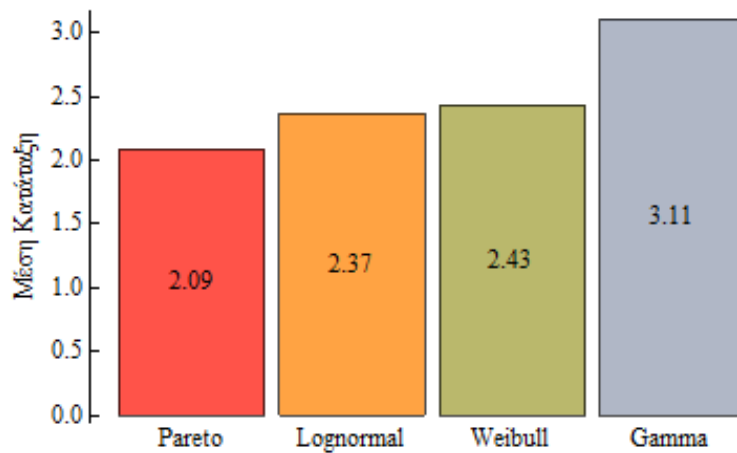
Οι τέσσερις τύποι ουρών προσαρμόζονται ακολουθώντας τη διαδικασία που περιγράφεται (βλ. Παράρτημα Β) στα δεδομένα των 3 477 σταθμών μέσω της ελαχιστοποίησης της νόρμας 4.17. Για να εκτιμηθεί ποια από τις τέσσερις κατανομές δίνει την καλύτερη εφαρμογή, οι τέσσερις ουρές συγκρίθηκαν σε ζεύγη βάσει του σφάλματος που έδωσαν, δηλαδή η ουρά με το μικρότερο σφάλμα θεωρήθηκε ότι έδινε την καλύτερη συγκριτικά εφαρμογή. Όπως φαίνεται στο Σχήμα 4.45, η κατανομή ουράς τύπου Pareto προσαρμόστηκε καλύτερα περίπου στο 60% των σταθμών, ενώ η πιο διαδεδομένη στις βροχοπτώσεις κατανομή, η Γάμμα, πάντα «ερχόταν δεύτερη» σε σχέση με την κατανομή που συγκρινόταν, με ποσοστό μόλις 30% των σταθμών.



Σχήμα 4.45 Σύγκριση των προσαρμοσμένων ουρών ανά δύο βάσει του Μέσου Τετραγωνικού Σφάλματος

Ενδιαφέρον παρουσιάζει το γεγονός ότι σε κάθε σύγκριση προέκυπτε πάντα καταλληλότερη στο μεγαλύτερο ποσοστό των σταθμών η πιο βαριά ουρά.

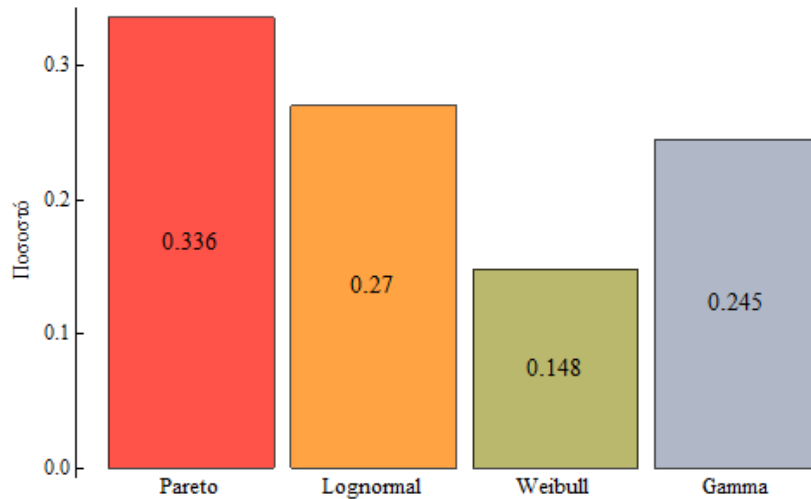
Επίσης πραγματοποιήθηκε κατάταξη της επίδοσης των τεσσάρων ουρών κατανομών μέσω βαθμονόμησης, ώστε να συγκριθούν συνολικά. Οι ουρές που εφαρμόστηκαν βαθμονομήθηκαν βάσει του σφάλματος, δηλαδή η ουρά με το μικρότερο σφάλμα (που είχε και την καλύτερη επίδοση) είχε βαθμό 1 ενώ αυτή με το μεγαλύτερο, βαθμό 4. Η μέση κατάταξη φαίνεται στο Σχήμα 4.46.



Σχήμα 4.46 Μέση βαθμονόμηση των ουρών που εφαρμόστηκαν για τα δεδομένα όλων των σταθμών. Το 1 αντιστοιχεί στην καλύτερη και το 4 στη χειρότερη εφαρμογή ουράς.

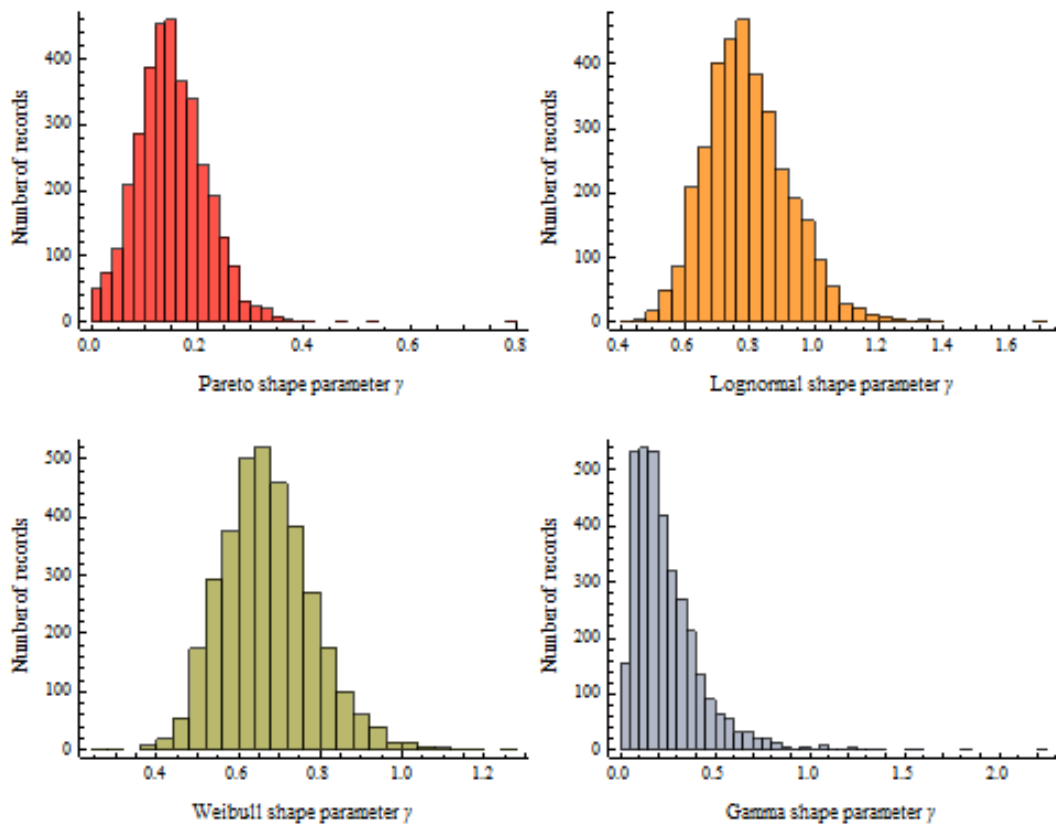
Και σε αυτή την περίπτωση η Pareto είχε την καλύτερη εφαρμογή ενώ η κατανομή Γάμμα είχε τη χειρότερη επίδοση.

Τέλος, τα ποσοστά για την καλύτερη εφαρμογή ουράς (Σχήμα 4.47) είναι 33.6% για την Pareto, 27% για τη Λογαριθμοκανονική, 14.8% για τη Weibull και 24.5% για τη Γάμμα. Η Γάμμα και σε αυτή την κατάταξη φαίνεται να είναι η καλύτερη. Το γεγονός ότι σ' αυτή την περίπτωση η Γάμμα έχει μεγαλύτερο ποσοστό από τη Weibull δεν έρχεται σε αντίθεση με τις παραπάνω κατατάξεις, καθώς, πιθανότατα, αν και η Γάμμα ήταν η καλύτερη επιλογή σε κάποιες περιπτώσεις, σε κάποιες άλλες ήταν η χειρότερη, όπως αποδεικνύεται από το Σχήμα 4.31.



Σχήμα 4.47 Ποσοστά καλύτερης εφαρμογής ουράς στο σύνολο των σταθμών.

Στο Σχήμα 4.48 φαίνονται οι εμπειρικές κατανομές των παραμέτρων σχήματος των κατανομών κατά την εφαρμογή τους στο σύνολο των σταθμών. Οι πιθανότερες τιμές όπως είναι γνωστό είναι εκεί που η συνάρτηση μάζας πιθανότητας παίρνει την υψηλότερη τιμή της. Για την κατανομή Pareto είναι η 0.142. Αυτό σημαίνει ότι οι δυναμικές κατανομές που ασυμπτωτικά συμπεριφέρονται ως Pareto δε θα έχουν πεπερασμένες ροπές δύναμης μεγαλύτερης από $1/0.142 \approx 7.1$. Το γεγονός μάλιστα ότι οι τιμές της παραμέτρου κοντά στο 0.2 είναι επίσης συνηθισμένες όπως φαίνεται στην εμπειρική κατανομή παραμέτρου σχήματος για την Pareto στο Σχήμα 4.31, σημαίνει ότι δε θα υπάρχουν ροπές 5^{ης} τάξης και άνω, σε αντίθεση με ό,τι αναφέρεται σε κάποιες δημοσιεύσεις. Η ουρά Weibull έχει πιθανότερη τιμή παραμέτρου σχήματος την $0.654 < 1$ που υποδηλώνει υποεκθετική ουρά. Τέλος, η Λογαριθμοκανονική ουρά έχει πιθανότερη τιμή παραμέτρου σχήματος 0.77 και η ουρά Γάμμα 0.144.



Σχήμα 4.48 Ιστογράμματα παραμέτρων σχήματος γ για όλες τις κατανομές ουρών που εφαρμόστηκαν

Τέλος, αντλήθηκαν από τα αποτελέσματα οι περιπτώσεις που καλύτερη εφαρμογή είχαν οι κατανομές ουρών με λεπτές ουρές, δηλαδή όλες οι κατανομές Γάμμα και όσες Weibull είχαν παράμετρο σχήματος μεγαλύτερη της μονάδας, με σκοπό τη σύγκριση με τα αποτελέσματα της Συνάρτησης Μέσης Υπέρβασης. Οι περιπτώσεις όπου η Γάμμα είχε την καλύτερη εφαρμογή, όπως έχει αναφερθεί, ανέρχονται στο 24,5% των δεδομένων ενώ μόνο σε 20 περιπτώσεις καλύτερη εφαρμογή είχε η κατανομή Weibull με παράμετρο σχήματος μεγαλύτερη της μονάδας. Άρα το τελικό ποσοστό προσαρμογής «εκθετικών» κατανομών από αυτή τη μέθοδο είναι 25%

5 ΣΥΝΟΨΗ ΚΑΙ ΣΥΜΠΕΡΑΣΜΑΤΑ

Από τις τρεις μεθόδους που αναλύθηκαν στο Κεφάλαιο 4 στο κεφάλαιο αυτό παρουσιάζονται τα βασικότερα αποτελέσματα και συμπεράσματα.

- Γραφική παράσταση σε λογαριθμικούς άξονες (Log-Log Plot)
 - Περιγραφή Μεθόδου

Η μέθοδος προτείνει την απεικόνιση σε λογαριθμικούς άξονες της συνάρτησης υπέρβασης συναρτήσεως των διατεταγμένων σε αύξουσα σειρά δεδομένων της ουράς. Στην περίπτωση που η γραφική παράσταση είναι ευθεία γραμμή, η κατανομή έχει χονδρή ουρά (ομαλά κυμαινόμενη κατανομή π.χ. Pareto). Για να γίνει δυνατή η εποπτία των δεδομένων και των 3 477 σταθμών, προσαρμόστηκε ευθεία στο γράφημα και υπολογίστηκε ο συντελεστής R^2 . Μεγάλες τιμές του συντελεστή υποδεικνύουν προσέγγιση σε ικανοποιητικό βαθμό ευθείας γραμμής.

- Εφαρμογή στα δεδομένα

Οι ουρές της πλειοψηφίας των σταθμών είχαν ιδιαίτερα υψηλές τιμές του συντελεστή R^2 . Συγκεκριμένα, ο μέσος όρος των τιμών του συντελεστή είναι 0.98 με τυπική απόκλιση 0.014 και η μέγιστη και ελάχιστη τιμή 0.997 και 0.857 αντίστοιχα.

- Προσομοίωση Monte Carlo

Πραγματοποιήθηκε προσομοίωση Monte Carlo σε θεωρητικές κατανομές ώστε να ελεγχθεί η αξιοπιστία των αποτελεσμάτων. Παρήχθησαν 1 000 τυχαία δείγματα από τις κατανομές προς εξέταση και συγκρίθηκαν τα αποτελέσματα. Το μέγεθος του δείγματος λήφθηκε ίσο με το μέσο όρο των μετρήσεων και ο αριθμός δεδομένων ουράς ίσος με το μέσο όρο των ετών των μετρήσεων όλων των σταθμών. Δεν ήταν όμως δυνατό να εξαχθούν συμπεράσματα, καθώς σε όλες τις κατανομές, είτε είχαν λεπτές είτε χονδρές ουρές, ο συντελεστής R^2 έπαιρνε πάντα υψηλές τιμές.

- Συμπέρασμα

Εφόσον ο συντελεστής R^2 , άρα και ο βαθμός που η συνάρτηση θεωρείται ευθεία, δε διαφοροποιείται κατά την εφαρμογή της μεθόδου στις θεωρητικές κατανομές, η μέθοδος κρίνεται αναξιόπιστη για την εξαγωγή συμπερασμάτων σχετικά με το είδος της ουράς.

- Συνάρτηση Μέσης Υπέρβασης (Mean Excess Function)

- Περιγραφή Μεθόδου

Η μέθοδος προτείνει τη σχεδίαση της εμπειρικής τιμής του $e(u)$ συναρτήσει του u , όπου $e(u)$ η συνάρτηση μέσης υπέρβασης για τα δεδομένα της ουράς. Στην περίπτωση που η γραφική παράσταση προκύπτει σταθερή (με μηδενική κλίση) θεωρείται ότι η κατανομή έχει εκθετική ουρά, λόγω της αμνήμονης ιδότητας της Εκθετικής κατανομής. Έτσι, για να πραγματοποιηθεί ενιαίος έλεγχος, στο κάθε γράφημα προσαρμόστηκε ευθεία της οποίας και βρέθηκε η κλίση.

- Προσομοίωση Monte Carlo

Πραγματοποιήθηκαν δοκιμές σε τυχαία δείγματα Εκθετικής κατανομής για να ελεγχθεί κατά πόσο επηρεάζονται οι κλίσεις του γραφήματος Σ.Μ.Υ. από την παράμετρο β της κατανομής καθώς και από το μέγεθος του συνόλου του δείγματος αλλά και το μέγεθος της ουράς. Προέκυψε ότι η παράμετρος και το μέγεθος του συνόλου του δείγματος δεν έχουν επίδραση στην κλίση, αντίθετα όμως αυτή επηρεάζεται από το μέγεθος του δείγματος ουράς, συναρτήσει του οποίου δημιουργήθηκε διάστημα εμπιστοσύνης των κλίσεων της Εκθετικής Κατανομής. Το διάστημα μάλιστα μοντελοποιήθηκε και δημιουργήθηκε τεστ το οποίο χαρακτηρίζει ως Εκθετική την κατανομή ουράς που δίνει κλίση που ανήκει μέσα στα όρια του διαστήματος, ως Υποεκθετική αυτήν της οποίας η κλίση είναι μεγαλύτερη από τα όρια και αντίστοιχα ως Υπερεκθετική αυτή με κλίση μικρότερη από τα όρια.

Επιπλέον, ελέγχθηκαν τα αποτελέσματα του τεστ για τις βασικές θεωρητικές κατανομές Pareto, Λογαριθμοκανονική, Weibull και Γάμμα. Από τις δοκιμές που έγιναν προέκυψε ότι η μέθοδος είναι πιο ισχυρή για μεγάλα δείγματα ουράς, ενώ παρουσιάζει μια αβεβαιότητα στα μικρά δείγματα.

- Εφαρμογή στα δεδομένα – Συμπεράσματα

Εφαρμόστηκε ο κατάλληλος αλγόριθμος στα δεδομένα (βλ. Παράρτημα Β) και βρέθηκε ότι αυτά έχουν κλίσεις γραφήματος Σ.Μ.Υ. με μέση τιμή 0.153, διάμεσο 0.138 και τυπική απόκλιση 0.097. Από την εφαρμογή του τεστ για διάστημα εμπιστοσύνης 90% προέκυψε ότι από το σύνολο των σταθμών ένα ποσοστό 33.5% είχε Σ.Μ.Υ. με κλίσεις μέσα στα όρια στα οποία ορίσαμε την Εκθετική Κατανομή, επομένως για αυτά τα δεδομένα δεν μπορούμε να

απορρίψουμε την υπόθεση της Εκθετικής ουράς. Το υπόλοιπο 66% θεωρήθηκε ότι έχει υποεκθετικού τύπου ουρές.

- Μέθοδος ελαχιστοποίησης σφάλματος
 - Επιλογή νόρμας σφάλματος

Για κάθε βασική θεωρητική κατανομή ουράς και για διάφορες παραμέτρους σχήματος παρήχθησαν 1 000 τυχαία δείγματα στα οποία εφαρμόστηκαν τέσσερις νόρμες που δίνουν το σφάλμα μεταξύ εμπειρικών δεδομένων και κατανομής που προσαρμόζεται. Στην προκειμένη περίπτωση τα εμπειρικά δεδομένα ήταν τυχαία παραγόμενα δείγματα από τις υπό εξέταση ουρές κατανομών για κάποιες τιμές παραμέτρων και η νόρμα ελαχιστοποιούταν για να προκύψουν παράμετροι ίσες με αυτές που θεωρήθηκαν αρχικά. Τελικά επιλέχτηκε η Σχέση 4.17 που έδινε τη μικρότερη απόκλιση παραμέτρου.

- Εφαρμογή στα δεδομένα – Αποτελέσματα

Οι τέσσερις τύποι ουρών προσαρμόζονται ακολουθώντας τη διαδικασία που περιγράφεται (βλ. Παράρτημα Β) στα δεδομένα των 3 477 σταθμών. Για να εκτιμηθεί ποια από τις τέσσερις κατανομές δίνει την καλύτερη εφαρμογή, οι τέσσερις ουρές συγκρίθηκαν σε ζεύγη βάσει του σφάλματος που έδωσαν. Η κατανομή ουράς τύπου Pareto προσαρμόστηκε καλύτερα περίπου στο 60% των σταθμών, ενώ η πιο διαδεδομένη στις βροχοπτώσεις κατανομή, η Γάμμα, πάντα «ερχόταν δεύτερη» σε σχέση με την κατανομή που συγκρινόταν, με ποσοστό μόλις 30% των σταθμών. Οι ουρές επίσης βαθμονομήθηκαν βάσει του σφάλματος, δηλαδή η ουρά με το μικρότερο σφάλμα (που είχε και την καλύτερη επίδοση) είχε βαθμό 1 ενώ αυτή με το μεγαλύτερο, βαθμό 4. Και πάλι η Pareto είχε την καλύτερη εφαρμογή ενώ η κατανομή Γάμμα είχε τη χειρότερη επίδοση. Τα ποσοστά για την καλύτερη εφαρμογή ουράς είναι 33.6% για την Pareto, 27% για τη Λογαριθμοκανονική, 14.8% για τη Weibull και 24.5% για τη Γάμμα.

- Γενικά Συμπεράσματα

- Η μέθοδος της Γραφικής Παράστασης σε Λογαριθμικούς άξονες δε συνιστά αξιόπιστο εργαλείο για το διαχωρισμό ουρών σε δυναμικές και μη.
- Η γραφική παράσταση της Συνάρτησης Μέσης Υπέρβασης μπορεί να θεωρηθεί ικανοποιητικός τρόπος κατάταξης ουρών σε Εκθετικές, Υπεοεκθετικές και Υπερεκθετικές για μεγάλα μήκη δείγματος, ενώ για μικρά μεροληπτεί υπέρ της επιλογής Εκθετικής ουράς. Αυτό οφείλεται στο ότι όσο πιο αραιά είναι τα διαθέσιμα δεδομένα, τόσο μεγαλύτερο είναι το διάστημα που κυμαίνονται οι κλίσεις της Σ.Μ.Υ. της Εκθετικής κατανομής, αφού χρειάζεται ικανοποιητικός αριθμός δεδομένων για να σταθεροποιηθεί και να φτάσει σε μηδενικές τιμές κατά τον ορισμό.
- Βάσει της μεθόδου Συνάρτησης Μέσης Υπέρβασης, το 33.7% του συνόλου των σταθμών έχει εκθετικές ουρές και το 66% υποεκθετικές.
- Βάσει της μεθόδου ελαχιστοποίησης σφάλματος, συνολικά οι εκθετικού τύπου κατανομές ουρών (κατανομή Γάμμα και κατανομή Weibull με παράμετρο σχήματος $\gamma > 1$) αποτελούν το 25% του συνόλου των σταθμών και το υπόλοιπο 75% περιγράφεται από χονδρές ουρές (Pareto, Λογαριθμοκανονική και Weibull με $\gamma < 1$).
- Οι δύο τελευταίες μέθοδοι συμφωνούν σε ικανοποιητικό βαθμό στην αναλογία εκθετικών και υποεκθετικών ουρών.
- Τα ακραία της βροχόπτωσης, λοιπόν, περιγράφονται καλύτερα από χονδρές ουρές.

6 ΒΙΒΛΙΟΓΡΑΦΙΚΕΣ ΑΝΑΦΟΡΕΣ

- Ahmad, M.I., C.D. Sinclair, and A. Werritty. 1988. Log-logistic Flood Frequency Analysis. *Journal of Hydrology* 98 (3–4) (April 15): 205–224.
- Balkema, A., & de Haan, L. 1974. Residual Life Time at Great Age. *The Annals of Probability* 2 (5) (October): 792–804.
- Benktander, G., & Segerdahl, C. 1960. On the analytical representation of claim distributions with special reference to excess of loss reinsurance. *in: XVIth International Congress of Actuaries*. Brussels.
- Bryson, M. C. (1974). Heavy-Tailed Distributions: Properties and Tests. *American Statistical Association and American Society for Quality*, Vol. 16, No. 1, 61-68.
- Buishand, T. (1978). Some remarks on the use of daily rainfall models. *J. Hydrol.*, 36, 295–308.
- Cline, D.B.H., and G. Samorodnitsky. 1994. Subexponentiality of the Product of Independent Random Variables. *Stochastic Processes and Their Applications* 49 (1) (January): 75–98.
- Cunnane, C. 1973. A Particular Comparison of Annual Maxima and Partial Duration Series Methods of Flood Frequency Prediction. *Journal of Hydrology* 18 (3–4) (March): 257–271.
- Dalle Molle, J. 2002. Modelling Catastrophic Loss Events Using Extreme Value Theory - Fundamental Results and Applications in Finance and Insurance. *New Zealand Society of Actuaries Conference*. Rotorua.
- El Adlouni, S., B. Bobée, and T.B.M.J. Ouarda. 2008. On the Tails of Extreme Event Distributions in Hydrology. *Journal of Hydrology* 355 (1–4) (June 20): 16–33. doi:10.1016/j.jhydrol.2008.02.011.
- Embrechts, Paul, Claudia Klüppelberg, and Thomas Mikosch. 1997. *Modelling Extremal Events: For Insurance and Finance*. Springer.
- Fama, E. F. 1965. Portfolio Analysis in a Stable Paretian Market. *Management Science*, (B2), 409–419

- Fisher, R., & Tippett, L. 1928. Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society*, Vol 24, 180-190.
- Gnedenko, B. 1941. Limit theorems for the maximal term of a variational series. *Doklady Akad. Nauk SSSR, Moscow*, 32, 37 (in Russian)
- Gnedenko, B. 1943. Sur La Distribution Limite Du Terme Maximum D'Une Serie Aleatoire. *The Annals of Mathematics* 44 (3) (July): 423
- Goldie, Charles M., and Claudia Klüppelberg. 1996. *Subexponential Distributions*. Johannes-Gutenberg-Univ.
- Gumbel, Emil Julius. 1958. *Statistics of extremes*. New York: Columbia University Press.
- Haan, Laurens de. 1976. Sample Extremes: An Elementary Introduction. *Statistica Neerlandica* 30 (4)
- Haan, Laurens de, and Ana Ferreira. 2006. *Extreme Value Theory: An Introduction*. 1st ed. Springer.
- Heo, J., Boes, D., & Salas, J. 2001a. Regional flood frequency analysis based on a Weibull model: Part 1. Estimation and asymptotic variances. *J. Hydrol*, 242, 157–170.
- Iii, James Pickands. 1975. Statistical Inference Using Extreme Order Statistics. *The Annals of Statistics* 3 (1) (January)
- Jenkinson, A. F. 1955. The Frequency Distribution of the Annual Maximum (or Minimum) Values of Meteorological Elements. *Quarterly Journal of the Royal Meteorological Society* 81 (348):
- Klüppelberg, Claudia, 1988. Subexponential Distributions and Integrated Tails. *Journal of Applied Probability* 25 (1) (March): 132.
- Koutsoyiannis, D. 2004. Statistics of Extremes and Estimation of Extreme Rainfall, 1, Theoretical Investigation. *Hydrological Sciences Journal* 49 (4): 575–590.
- Leadbetter, M. R. 1974. On Extreme Values in Stationary Sequences. *Probability Theory and Related Fields* 28 (4): 289–303.

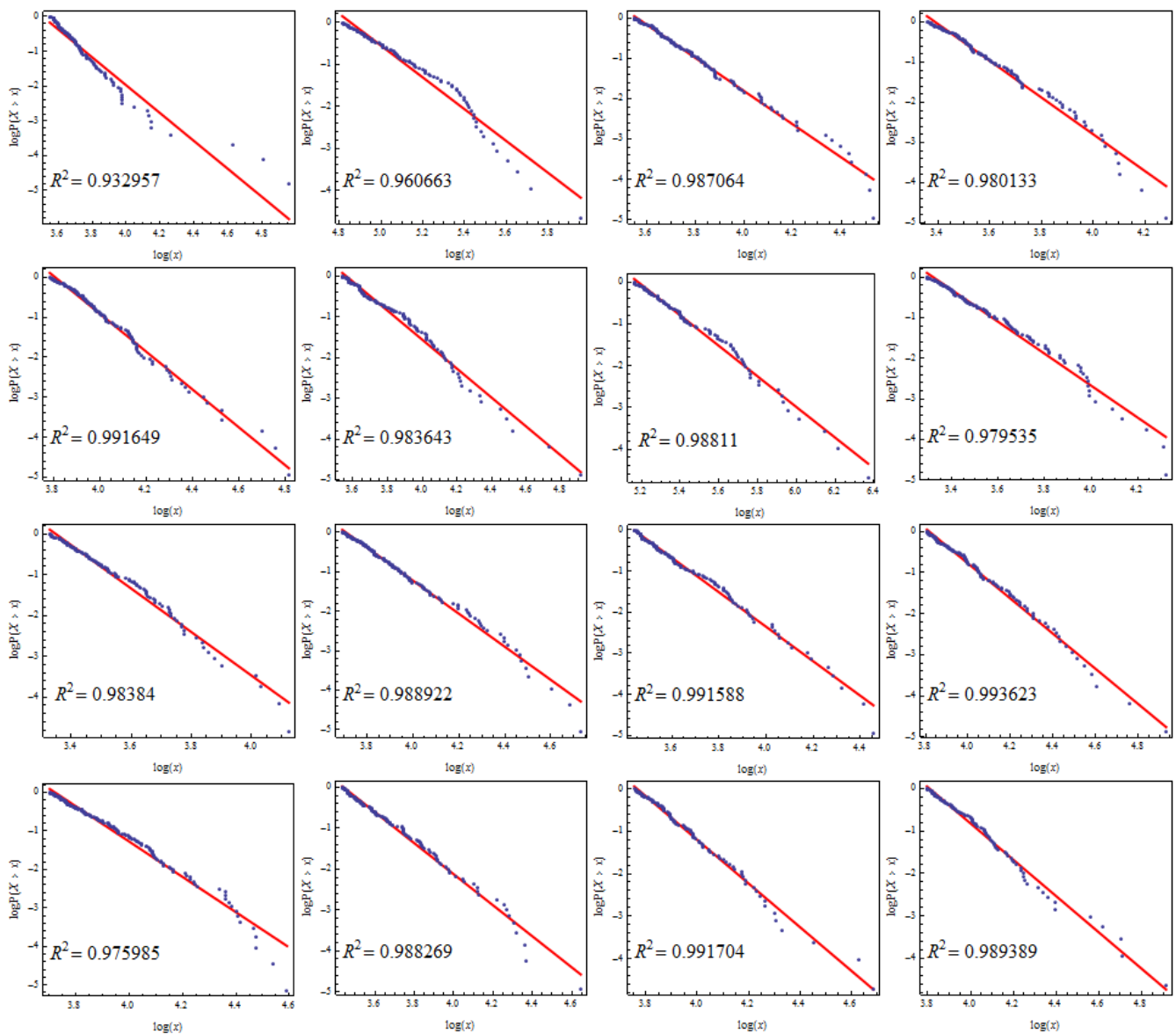
- Makkonen, Lasse. 2006. "Plotting Positions in Extreme Value Analysis." *Journal of Applied Meteorology and Climatology* 45 (2) (February): 334–340.
- Mandelbrot, B. 2003. Multifractal power law distributions: negative and critical dimensions and other "anomalies", explained by a simple example. *Journal of Statistical Physics*, 110 (3–6), 739–774.
- McNeil, A. 1998. *Calculating Quantile Risk Measures for Financial Return Series*. Zurich, Switzerland: Department of Mathematics. ETH.
- McNeil, Alexander J., Rüdiger Frey, and Paul Embrechts. 2005. *Quantitative Risk Management: Concepts, Techniques, and Tools*. Princeton University Press.
- Mielke, Paul W. 1973. "Another Family of Distributions for Describing and Analyzing Precipitation Data." *Journal of Applied Meteorology* 12 (2) (March): 275–280.
- Mielke, Jr, Paul W. and Earl S. Johnson. 1974. "Some generalized beta distributions of the second kind having desirable application features in hydrology and meteorology." *Water Resources Research* 10 (2): 223–226.
- Mitzenmacher, M. (2004). A brief history of generative models for power law and lognormal distributions. *Internet Math*, 1, 226–251.
- Papalexiou, S.M., D. Koutsoyiannis, and C. Makropoulos. 2012. How Extreme Is Extreme? An Assessment of Daily Rainfall Distribution Tails. *Hydrology and Earth System Sciences Discussions* 9: 5757–5778.
- Pickands, J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics*, 3, 119–131.
- Reiss, R. D., & Thomas, M. (2007). *Statistical Analysis of Extreme Values*. Birkhäuser.
- Rossi, F., Fiorentino, M., & Versace, P. (1984). Two-component extreme value distribution for flood frequency analysis. *Water Resour. Res.*, 20(7), 847–856.
- Stedinger, J. R., Vogel, R. M., & Foufoula-Georgiou, E. (1993). Frequency Analysis of Extreme Events, Chapter 18 in *Handbook of Hydrology*. Maidment, DR, McGraw-Hill.
- Tadikamalla, Pandu R. 1980. A Look at the Burr and Related Distributions. *International Statistical Review / Revue Internationale De Statistique* 48 (3) (December): 337.

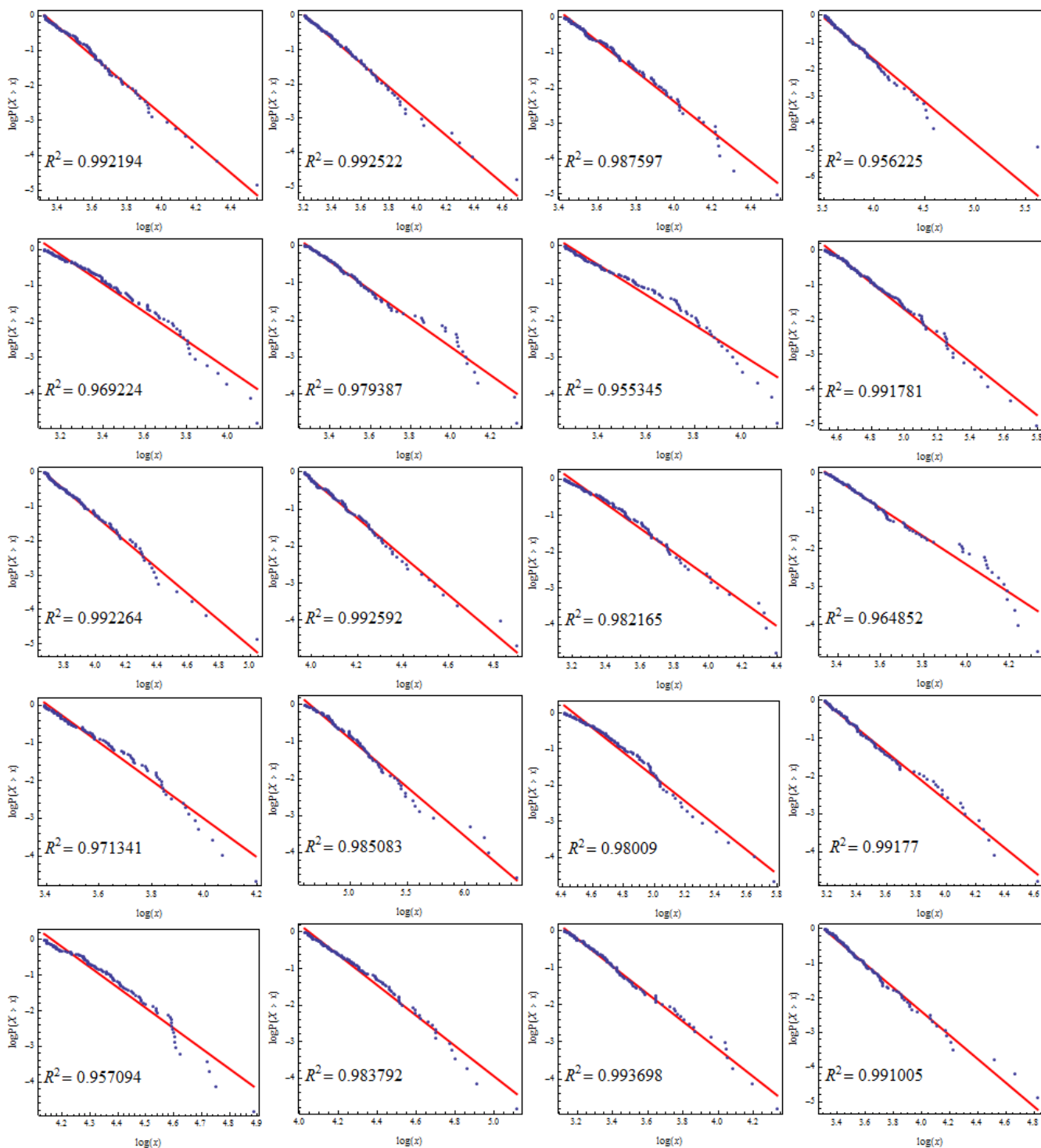
- Tavares, L.Valadares, and J.Evaristo Da Silva. 1983. Partial Duration Series Method Revisited. *Journal of Hydrology* 64 (1–4) (July): 1–14.
- Teugels, Jozef L. 1975. The Class of Subexponential Distributions. *The Annals of Probability* 3 (6) (December): 1000–1011.
- Todorovic, P., and E. Zelenhasic. 1970. A Stochastic Model for Flood Analysis. *Water Resources Research* 6 (6): 1641.
- von Mises, R. (1936). La Distribution de la plus Grande de n Valeurs. *Rev. of Math. Union Interbalcanique*, Vol. 1, 141-160.
- Weibull, Waloddi. 1939. *A Statistical Theory of the Strength of Materials*. Generalstabens litografiska anstalts förlag.
- Werner, Thomas, and Christian Upper. 2004. Time Variation in the Tail Behavior of Bund Future Returns. *Journal of Futures Markets* 24 (4): 387–398. doi:10.1002/fut.10120.
- Ευστρατιάδης, Α. 2011. Σημειώσεις υδρολογίας πλημμυρών και σχεδιασμού έργων αποχέτευσης ομβρίων. <http://www.itia.ntua.gr/el/docinfo/1154/>.

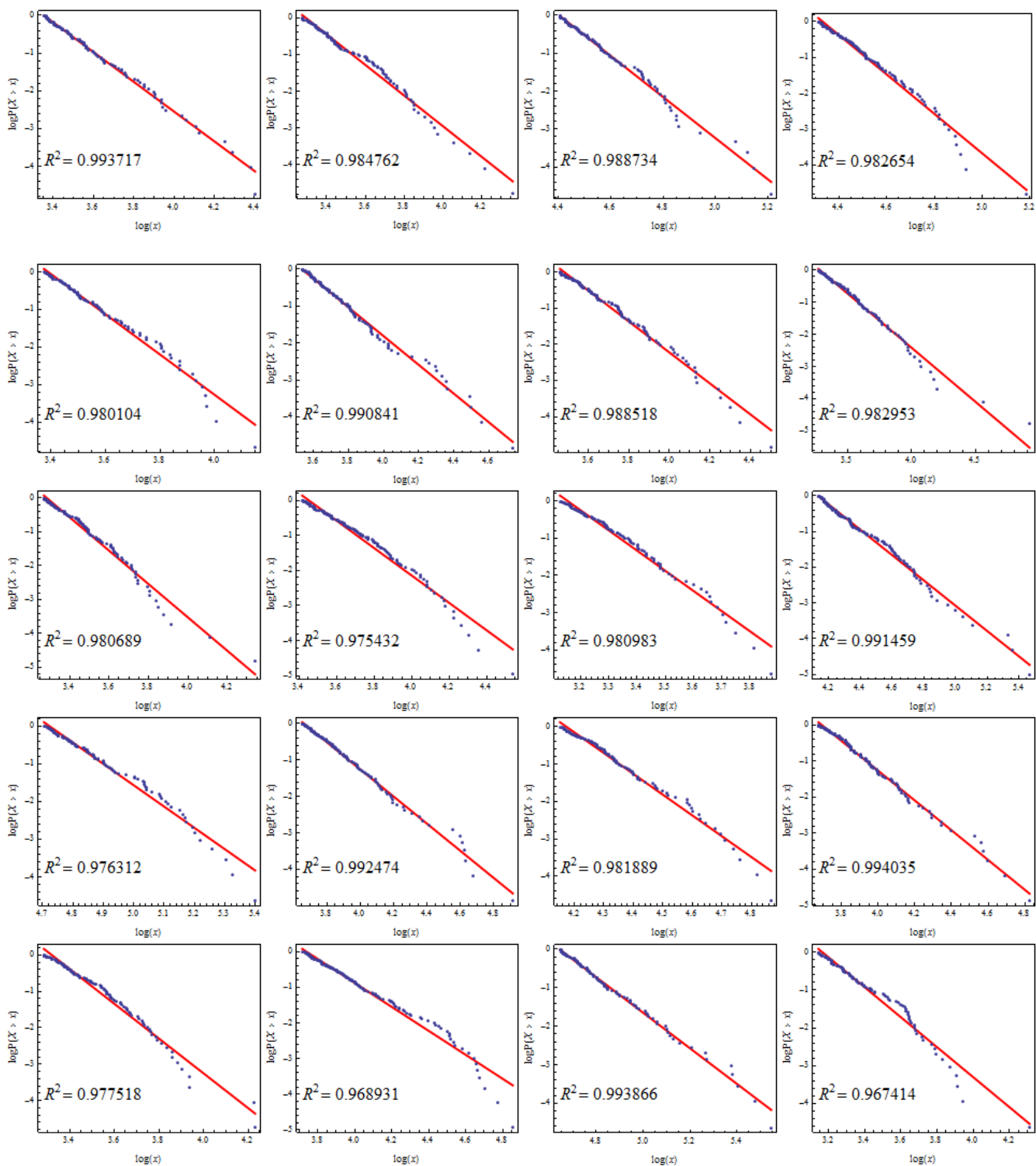
ΠΑΡΑΡΤΗΜΑ Α

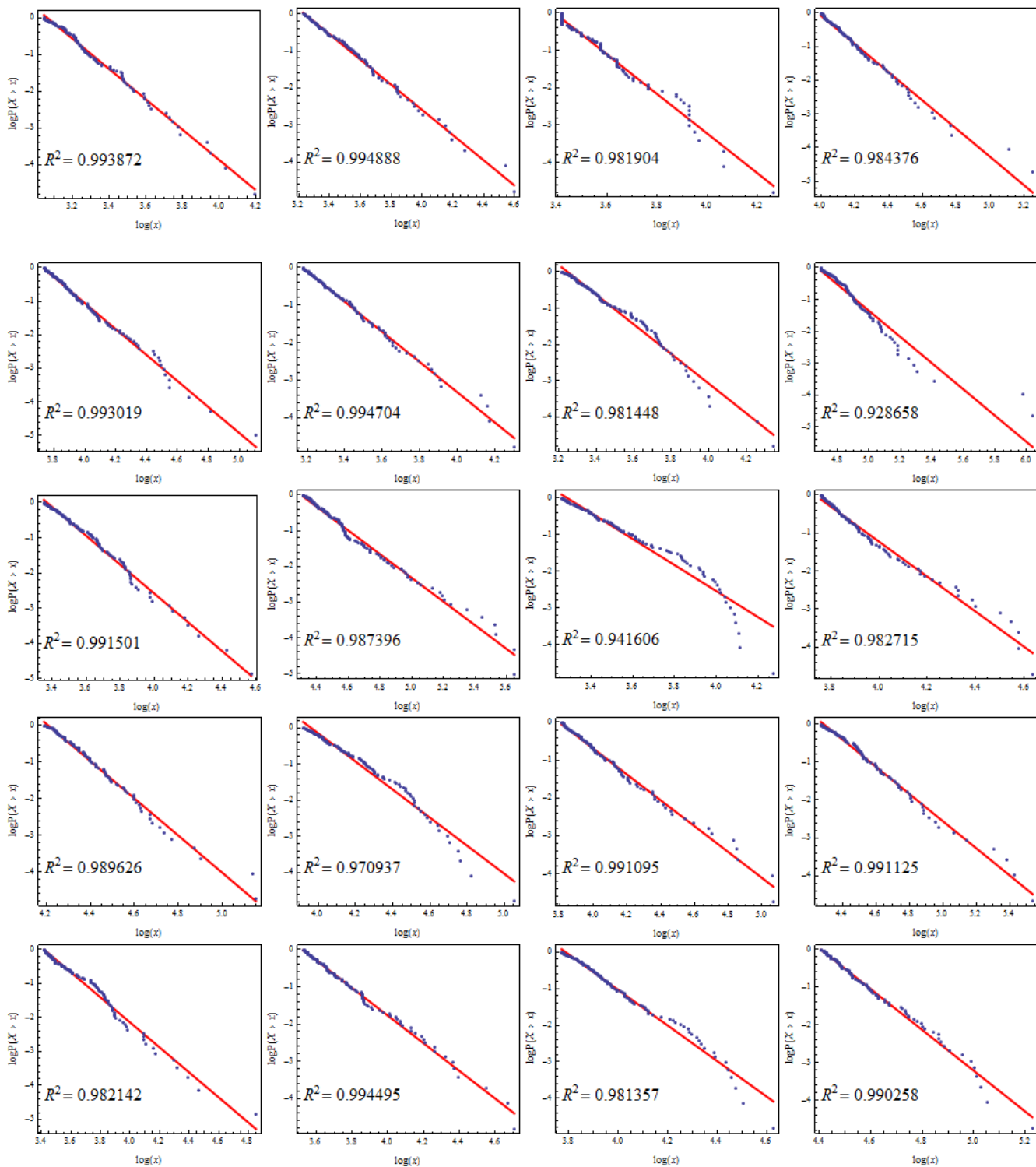
Α.1 ΓΡΑΦΗΜΑΤΑ ΣΕ ΛΟΓΑΡΙΘΜΙΚΟΥΣ ΑΞΟΝΕΣ

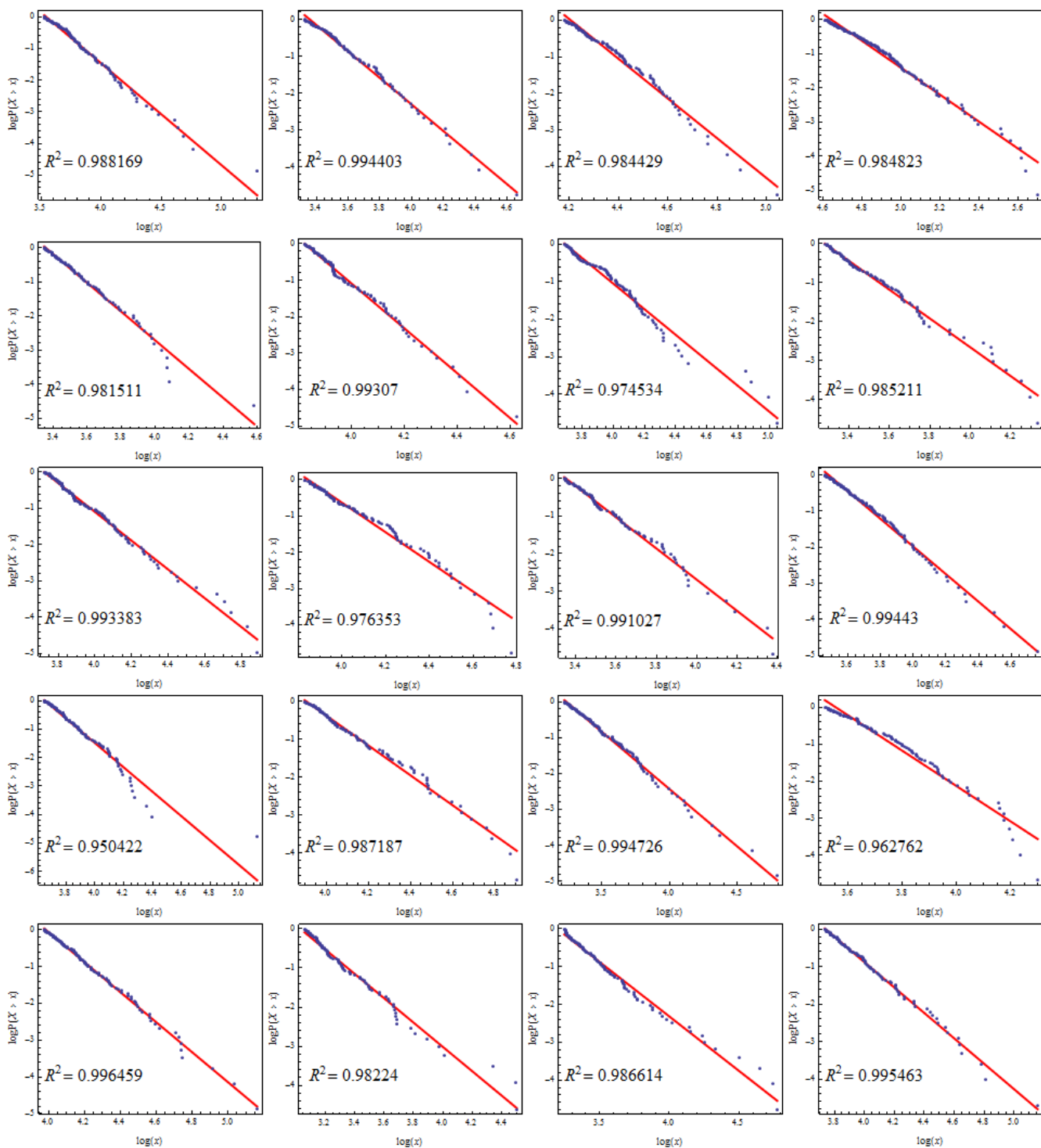
Για να γίνει ευκρινέστερη η αντιστοιχία συντελεστή R^2 και «γραμμικότητας», παρατίθενται διαγράμματα για τους 100 σταθμούς με τις περισσότερες μετρήσεις (Σχήμα Α.1).

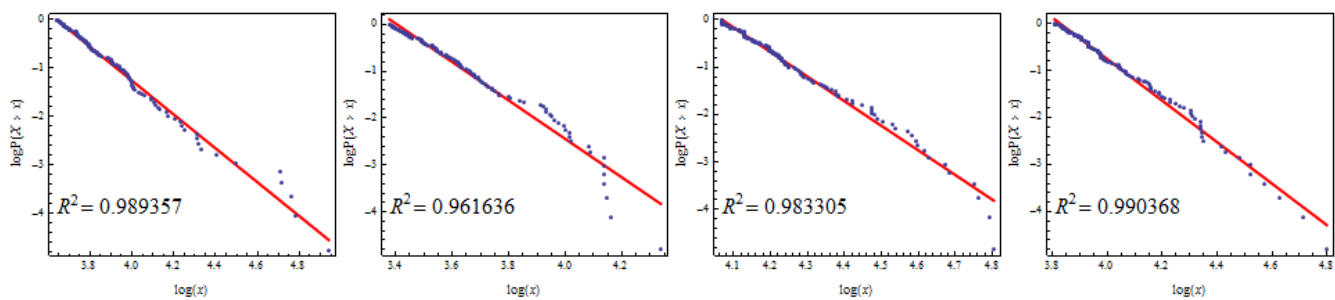








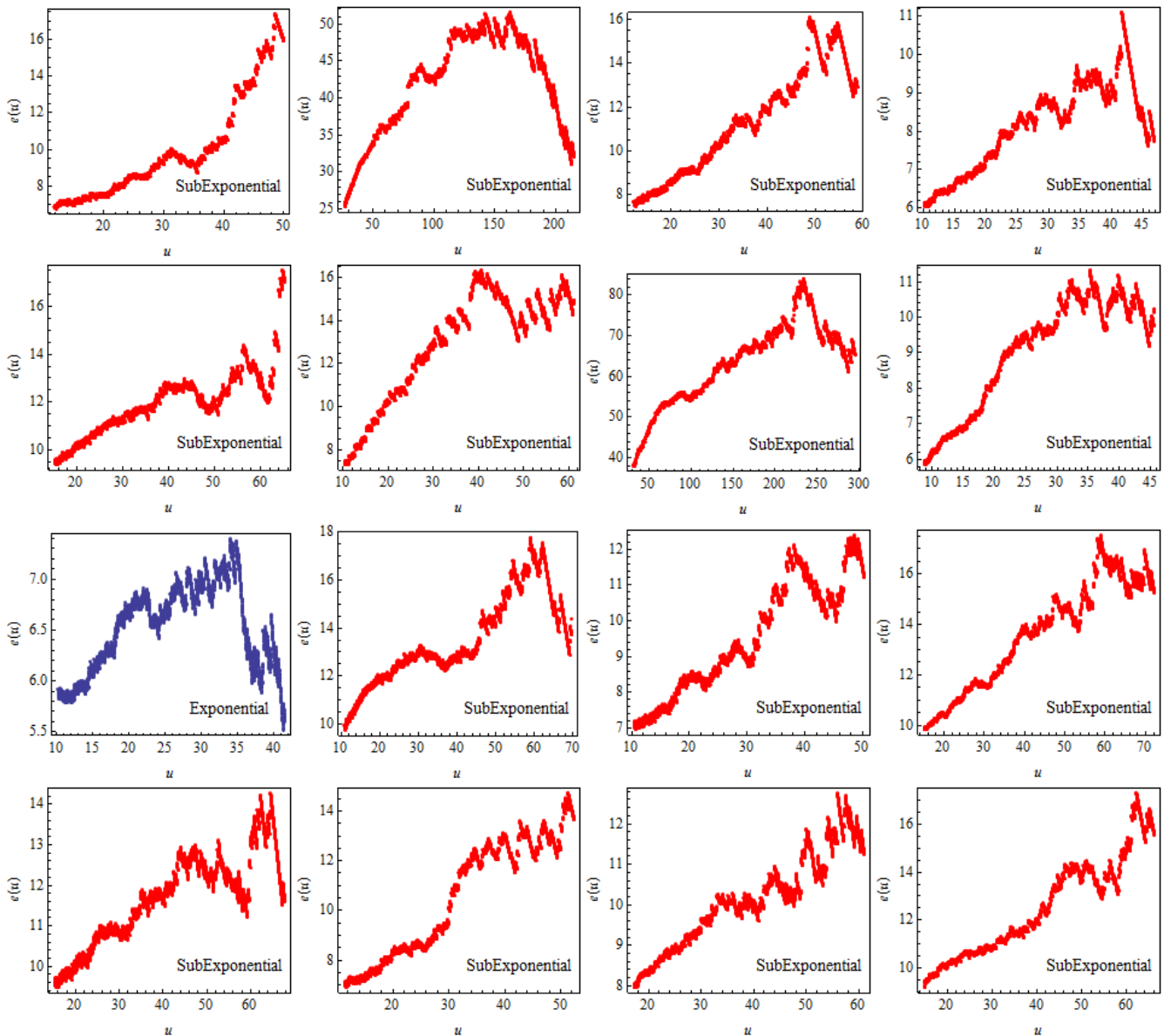


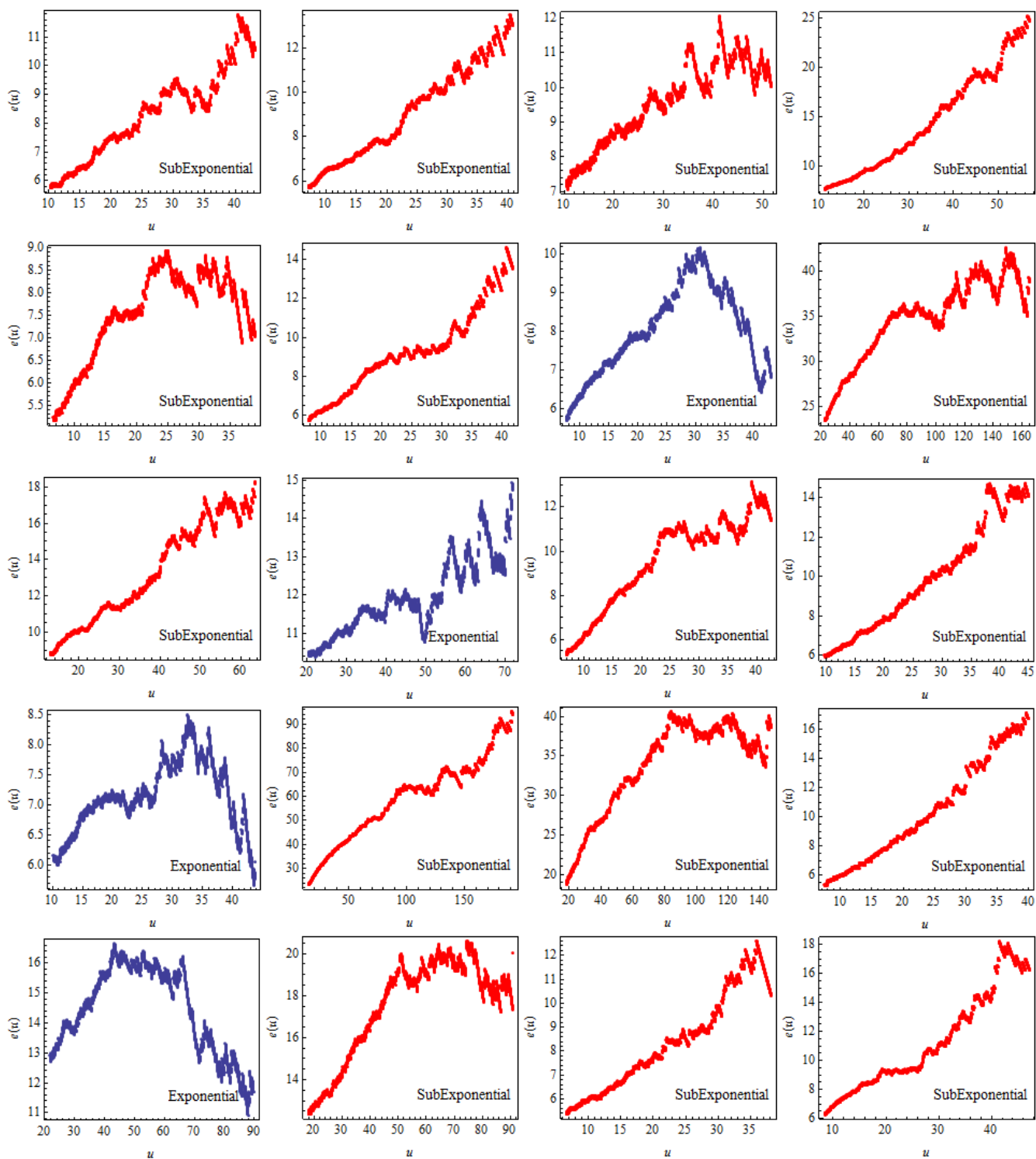


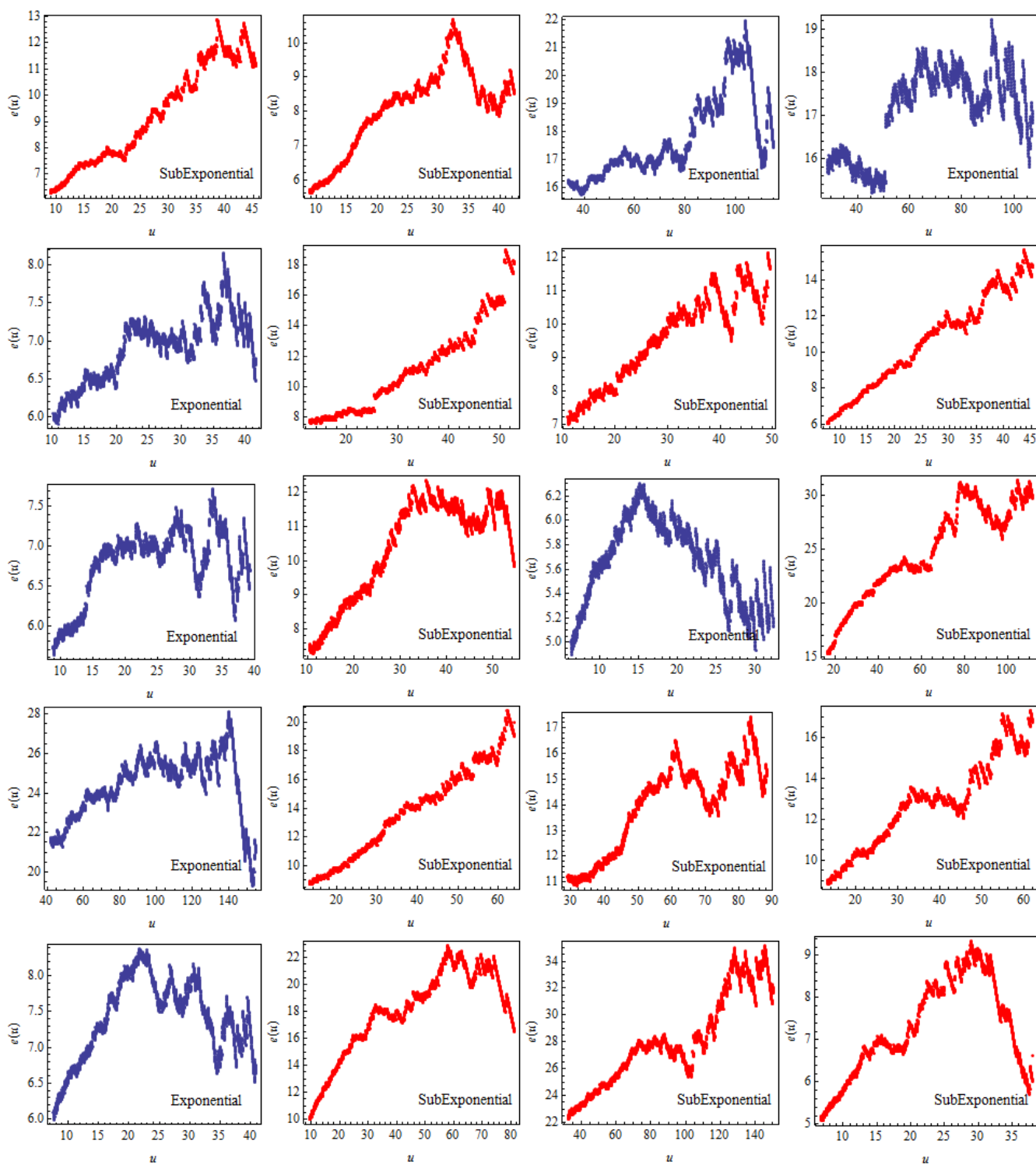
Σχήμα Α.1 Διπλά λογαριθμισμένη ουρά εμπειρικών τιμών και συντελεστής R^2 για τους 100 σταθμούς με τα περισσότερα δεδομένα.

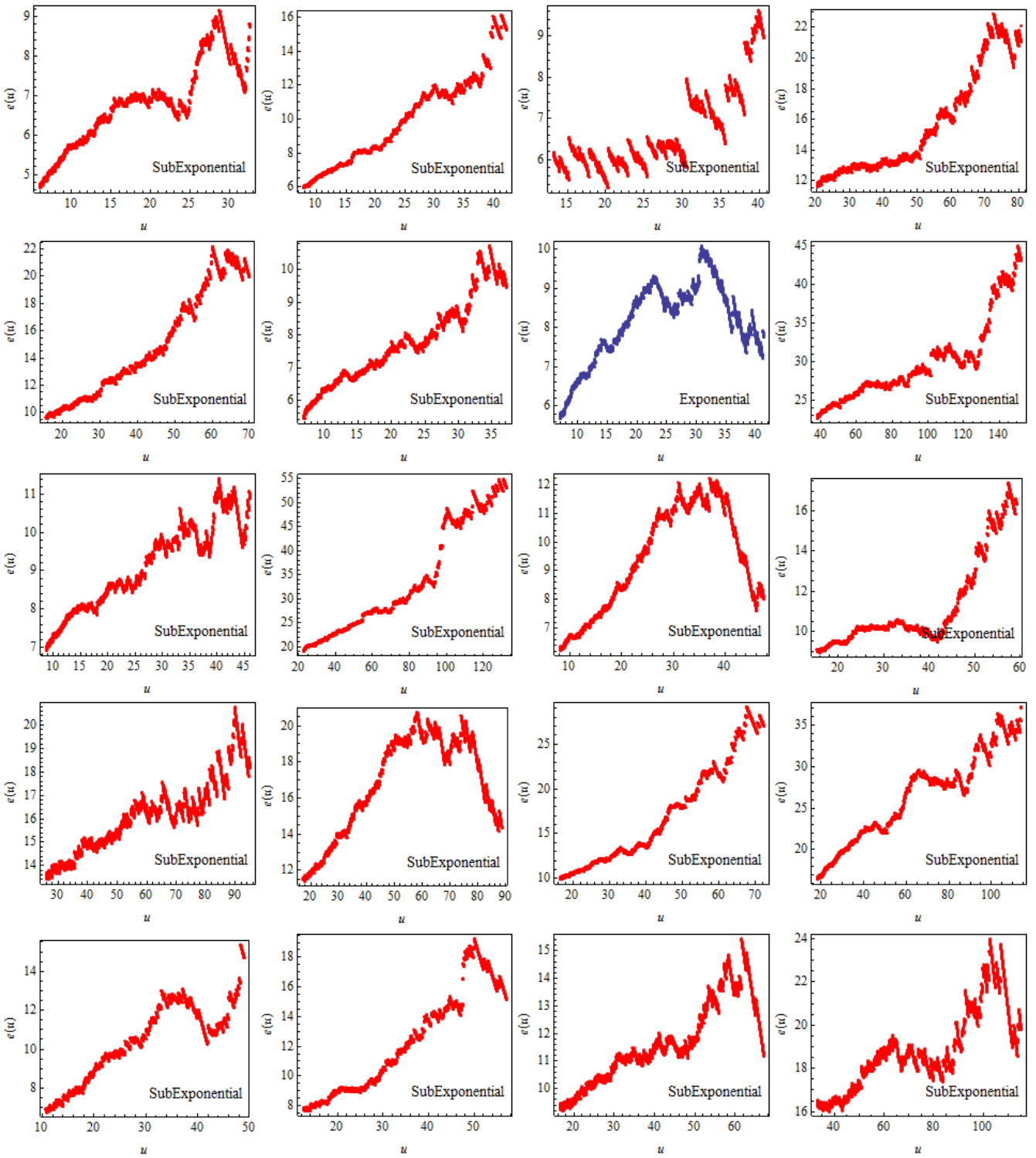
Α.2 ΓΡΑΦΗΜΑΤΑ ΣΥΝΑΡΤΗΣΗΣ ΜΕΣΗΣ ΥΠΕΡΒΑΣΗΣ

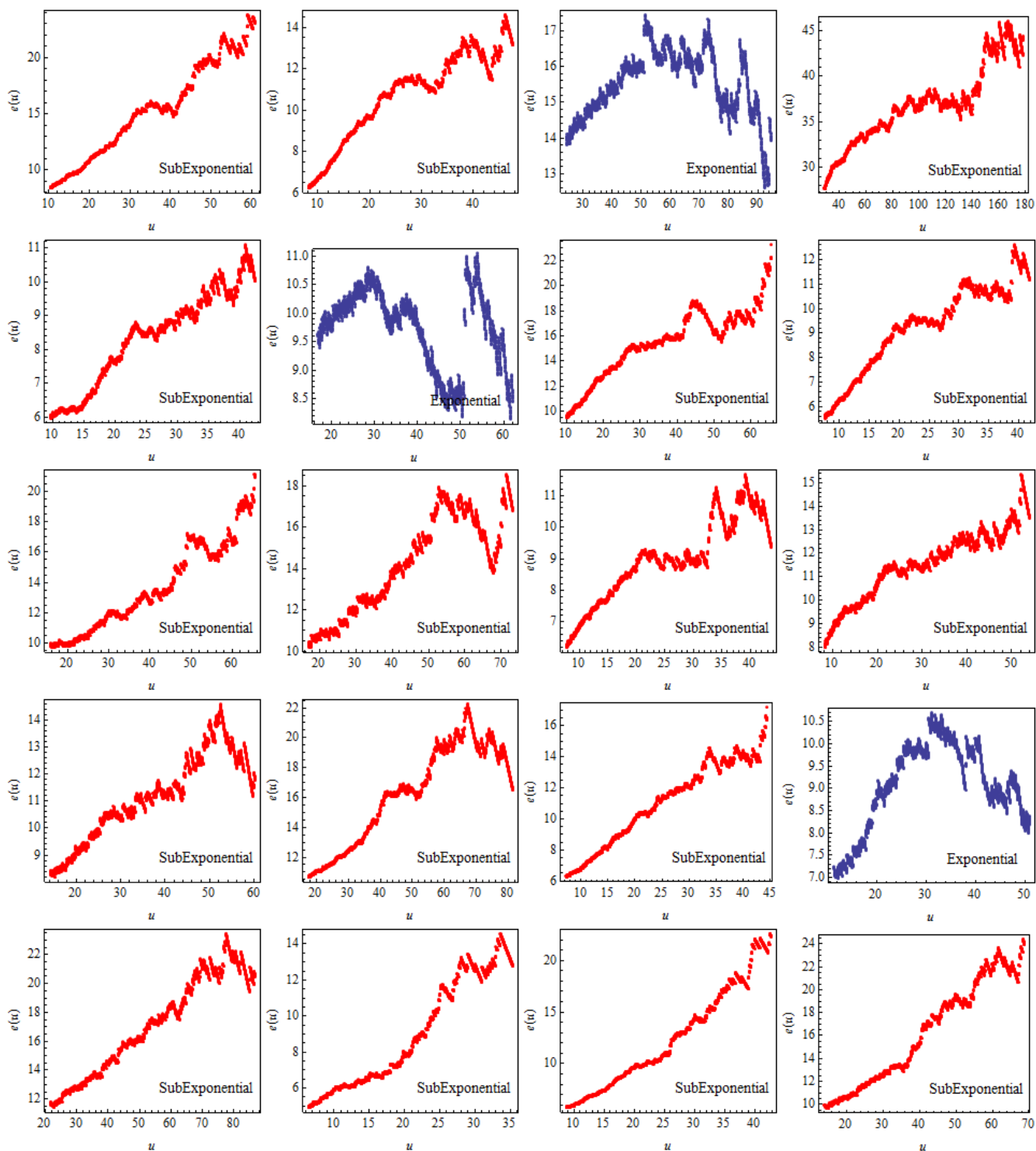
Για τους 100 σταθμούς με τις περισσότερες μετρήσεις ακολουθούν οι γραφικές παραστάσεις της Σ.Μ.Υ. τους. Αυτές με κόκκινο χρώμα υποδεικνύουν υποεκθετική ουρά ενώ με μπλε εκθετική (Σχήμα Α.2).

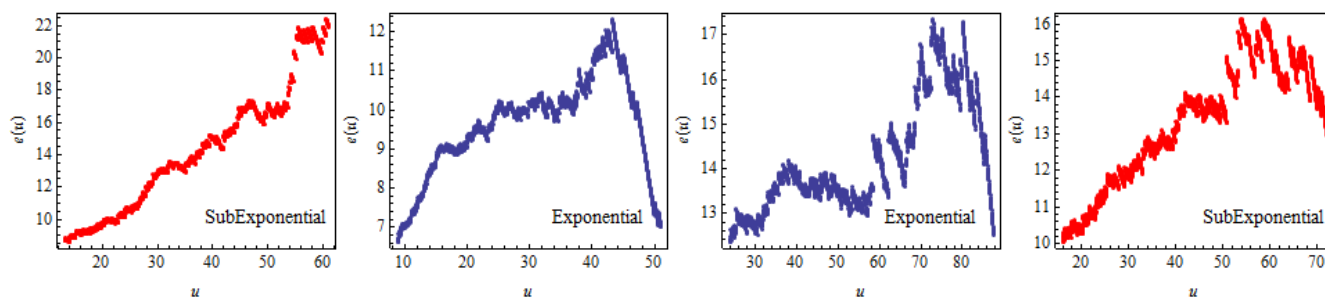










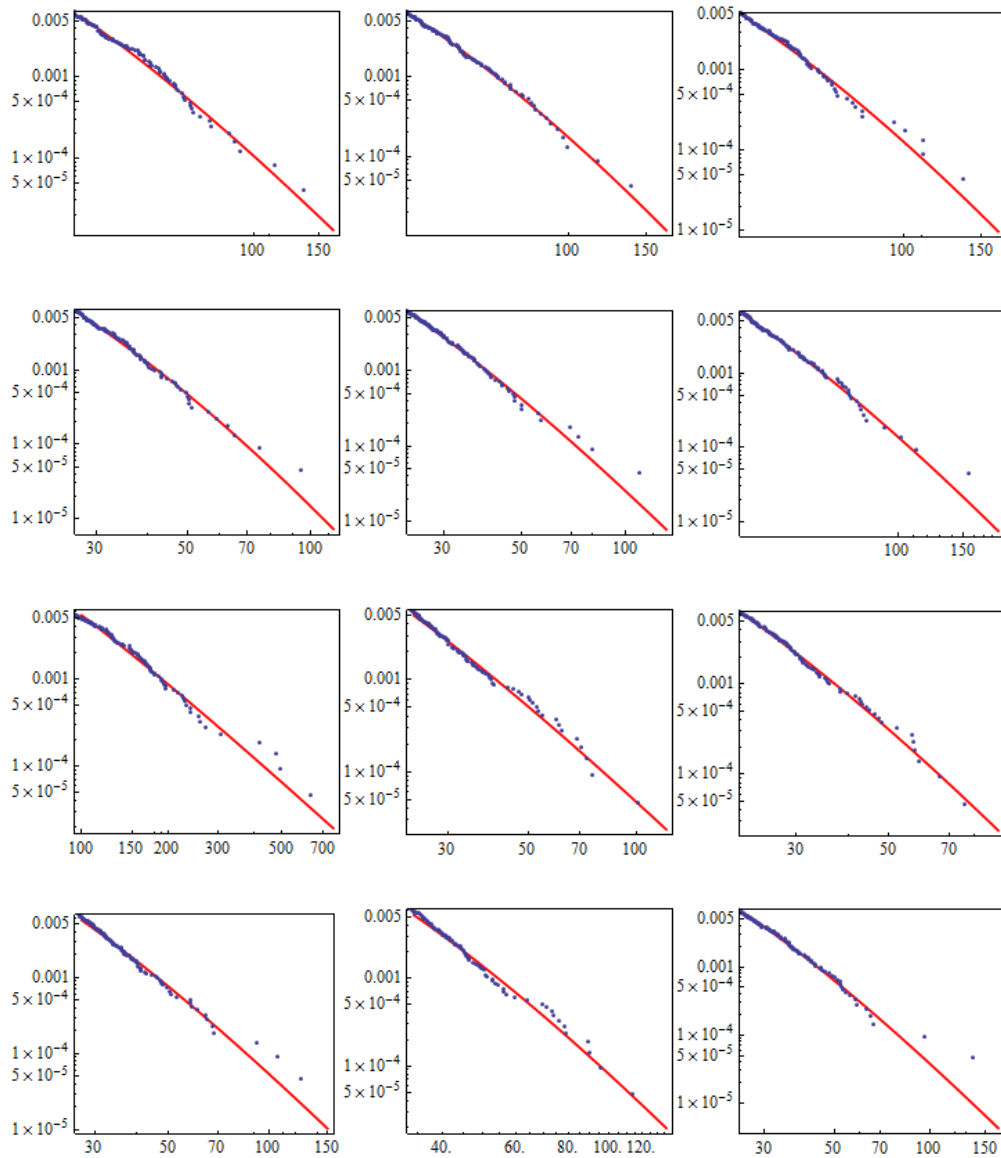


Σχήμα Α.2 Γραφικές παραστάσεις συνάρτησης μέσης υπέρβασης για τους 100 σταθμούς με τις περισσότερες παρατηρήσεις

Α.3 ΓΡΑΦΗΜΑΤΑ ΕΦΑΡΜΟΓΗΣ ΚΑΤΑΝΟΜΩΝ ΣΕ ΛΟΓΑΡΙΘΜΙΚΟΥΣ ΑΞΟΝΕΣ

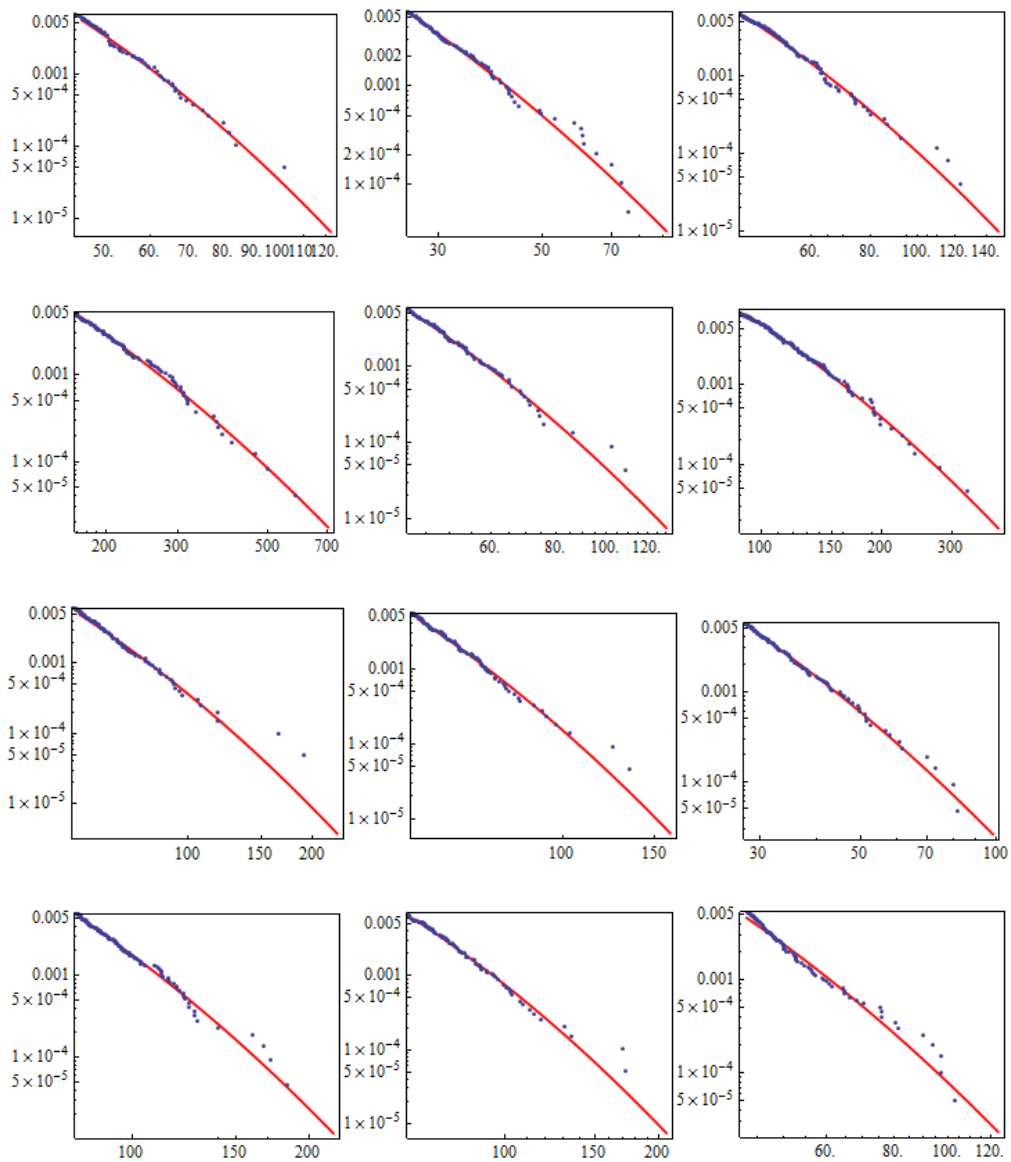
Παρατίθενται γραφήματα που δείχνουν την εφαρμογή της καταλληλότερης κατανομής για τα δεδομένα της ουράς των σταθμών με τις περισσότερες παρατηρήσεις.

Α.3.1 Κατανομή ουράς Pareto τύπου II



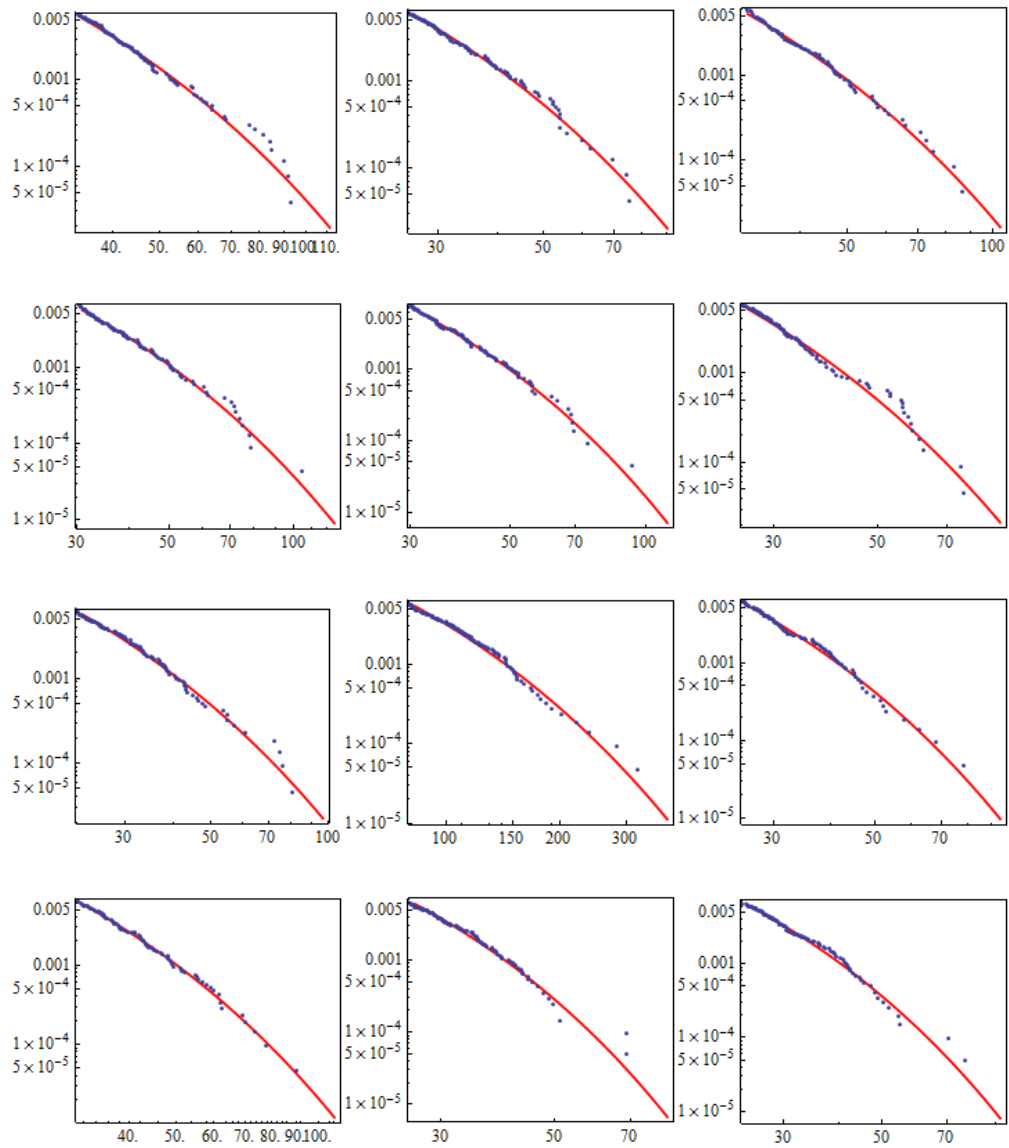
Σχήμα Α.3 Γραφικές παραστάσεις εμπειρικών δεδομένων ουράς και προσαρμοσμένης κατανομής ουράς Pareto σε λογαριθμικούς άξονες.

A.3.2 Λογαριθμοκανονική κατανομή ουράς



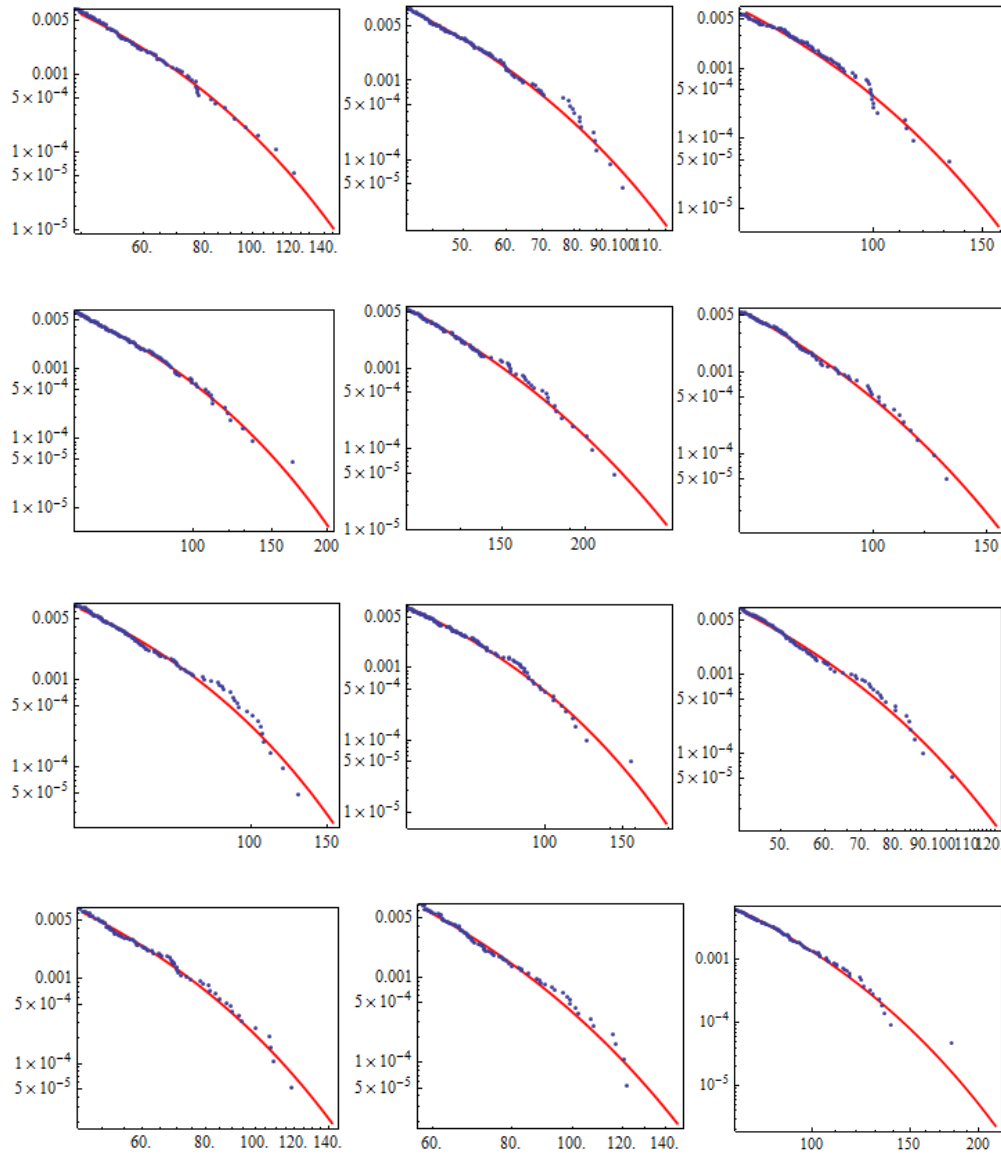
Σχήμα A.4 Γραφικές παραστάσεις εμπειρικών δεδομένων ουράς και προσαρμοσμένης Λογαριθμοκανονικής κατανομής ουράς σε λογαριθμικούς άξονες.

A.3.3 Κατανομή ουράς Weibull



Σχήμα A.5 Γραφικές παραστάσεις εμπειρικών δεδομένων ουράς και προσαρμοσμένης κατανομής ουράς Weibull σε λογαριθμικούς άξονες.

A.3.4 Κατανομή ουράς Γάμμα



Σχήμα Α.6 Γραφικές παραστάσεις εμπειρικών δεδομένων ουράς και προσαρμοσμένης κατανομής ουράς Γάμμα σε λογαριθμικούς άξονες

ΠΑΡΑΡΤΗΜΑ Β

ΠΡΟΓΡΑΜΜΑΤΙΣΜΟΣ

Η επεξεργασία των δεδομένων έγινε με χρήση του προγράμματος Wolfram Mathematica 8.

- Συνάρτηση Μέσης Υπέρβασης

Στη μέθοδο της Συνάρτησης Μέσης Υπέρβασης ακολουθήθηκε ο εξής αλγόριθμος:

```
MEFList[data_, minNoOfVals_] := Block[{TailData, TailMin, TailMax, Step},
  TailData = Take[Sort[data], -Round[0.1 * Length[data], 1]];
  TailMin = Min[TailData];
  TailMax = RankedMax[TailData, minNoOfVals];
  Step =  $\frac{\text{TailMax} - \text{TailMin}}{0.1 * \text{Length}[data]}$ ;
  Table[{u, Mean[Cases[TailData - u, _?Positive]]}, {u, TailMin, TailMax, Step}]];
```

Αρχικά ορίζονται τα δεδομένα ουράς *TailData* ως το τελευταίο 10% των διατεταγμένων εμπειρικών δεδομένων. Θεωρείται *TailMin* η ελάχιστη τιμή των δεδομένων ουράς και *TailMax* η 20^η μεγαλύτερη τιμή έτσι ώστε να υπάρχουν 20 τιμές μεγαλύτερες από τη μεγαλύτερη παρατήρηση. Λαμβάνεται βήμα *Step* για τις τιμές του *u* από *TailMin* ως *TailMax* και η Σ.Μ.Υ. ορίζεται ως ο μέσος όρος της διαφοράς των τιμών με το *u*, με τη δέσμευση αυτή να είναι θετική.

- Τεστ «Εκθετικότητας»

Για την περίπτωση διαστήματος εμπιστοσύνης 90%, εφαρμόζουμε το τεστ:

```
test90 = Table[

slope = Fit[MEFList[data[StNo], 20], {1, x}, x][[2, 1]];

Which[slope < -  $\frac{0.9292}{\text{Log}[1 + 0.1399 * \text{Round}[0.1 * \text{Length}[\text{data}[\text{StNo}], 1]]^{1.4957}}$ , "HyperExponential",
-  $\frac{0.9292}{\text{Log}[1 + 0.1399 * \text{Round}[0.1 * \text{Length}[\text{data}[\text{StNo}], 1]]^{1.4957}}$  ≤ slope ≤
 $\frac{0.4140}{\text{Log}[1 + 0.0419 * \text{Round}[0.1 * \text{Length}[\text{data}[\text{StNo}], 1]]^{1.2531}}$ , "Exponential",
slope >  $\frac{0.4140}{\text{Log}[1 + 0.0419 * \text{Round}[0.1 * \text{Length}[\text{data}[\text{StNo}], 1]]^{1.2531}}$ , "SubExponential"],
{StNo, 1, 3477}]
```

στο οποίο ορίζουμε ως *slope* την κλίση του γραφήματος Σ.Μ.Υ. των δεδομένων κάθε σταθμού και *n* το μήκος των δεδομένων. Εάν η κλίση προκύψει μικρότερη από τις τιμές της καμπύλης του μοντέλου που προσαρμόζει τα κάτω όρια των τιμών, τότε τα δεδομένα ενδέχεται να προσαρμόζονται καλύτερα από μια υπερεκθετική κατανομή. Εάν, αντίθετα, η κλίση προκύψει μεγαλύτερη από τα άνω όρια, είμαστε στην περίπτωση της υποεκθετικής κατανομής. Εάν, τέλος, προκύψει μέσα στα όρια που ορίζονται από την καμπύλη του μοντέλου, τότε είναι πιθανό τα δεδομένα να προσαρμόζονται από την Εκθετική Κατανομή.

- Σύγκριση τύπων σφάλματος

```

failExpr = {Null, {b → Null, c → Null}};

n = 116;
m = 9500;

DST = ParetoIIDistribution[b, c];
data := Quantile[DST, RandomReal[1, m]] /. {b → 10, c → 0.2};

ParallelTable[

  empDistPnts = EmpDistribution[data];

  EmpTailPr = Take[empDistPnts[[All, 2]], -n];
  EmpTailVals = Take[empDistPnts[[All, 1]], -n];

  ThTailPr = Take[1 - CDF[DST, EmpTailVals], -n];
  ThTailVals = Take[Quantile[DST, 1 - EmpTailPr], -n];

  M1 = Total[(ThTailVals - EmpTailVals)2];
  M2 = Total[( $\frac{\text{ThTailVals}}{\text{EmpTailVals}} - 1$ )2];
  M3 = Total[(ThTailPr - EmpTailPr)2];
  M4 = Total[( $\frac{\text{ThTailPr}}{\text{EmpTailPr}} - 1$ )2];

  cons = {5 < b < 15, 0.001 < c < .5};

  fit = {
    TimeConstrained[Check[NMinimize[{M1, cons}, {b, c}, AccuracyGoal → 3,
      PrecisionGoal → 3, MaxIterations → 100], failExpr], 100, failExpr],
    TimeConstrained[Check[NMinimize[{M2, cons}, {b, c}, AccuracyGoal → 3,
      PrecisionGoal → 3, MaxIterations → 100], failExpr], 100, failExpr],
    TimeConstrained[Check[NMinimize[{M3, cons}, {b, c}, AccuracyGoal → 3,
      PrecisionGoal → 3, MaxIterations → 100], failExpr], 100, failExpr],
    TimeConstrained[Check[NMinimize[{M4, cons}, {b, c}, AccuracyGoal → 3,
      PrecisionGoal → 3, MaxIterations → 100], failExpr], 100, failExpr]
  ];

  Export[StringTake[ToString[PaddedForm[i, 5, NumberPadding → "0"]], -4], fit, "List"],

  {i, 1000}];

```

Έστω N ο μέσος όρος του πλήθους των ετών και n ο μέσος όρος του πλήθους των παρατηρήσεων. Η DST είναι η θεωρητική κατανομή με παραμέτρους c και b . Ως $data$ συμβολίζονται οι τυχαίες τιμές της θεωρητικής κατανομής (n στον αριθμό). Οι συντεταγμένες της εμπειρικής κατανομής $empDistPnts$ παράγονται από την εμπειρική κατανομή $EmpDistribution$ των $data$, η οποία αντιστοιχίζει τις τυχαίες μεταβλητές με τη σχεδιαστική τους θέση (κατά Weibull). Έτσι, η $EmpTailPr$ είναι η πιθανότητα υπέρβασης της εμπειρικής κατανομής και $EmpTailVals$ είναι οι τιμές της εμπειρικής κατανομής. Όσον αφορά στη θεωρητική κατανομή, η $ThTailPr$ είναι η θεωρητική πιθανότητα υπέρβασης που ισούται με τη συμπληρωματική της αθροιστικής συνάρτησης πιθανότητας για τις τιμές της εμπειρικής $EmpTailVals$, ενώ οι θεωρητικές τιμές είναι το $1 - EmpTailPr$ ποσοστημόριο της θεωρητικής κατανομής.

- Εφαρμογή στα δεδομένα

```
FitDistributionTail[data_, pntsNo_, "Pareto", OptionsPattern[]] :=

Block[{xMin, tailData, n, model, parsFirstEstimate, b, c, cons, pars, ER, OF},

tailData = If[pntsNo == "All", EmpDistribution[data], xMin = Sort[data][[-pntsNo]];
Cases[EmpDistribution[data], {val_ /; val >= xMin, _}]];
n = Length[tailData];

cons = {0.01 < b < 75, 0.001 < c < 1};
ER[{x_, F1_}] := ((1 + c x / b) ^ - (1 / c) / F1 - 1) ^ 2;
OF = Total[ER /@ tailData] / n;

Flatten[NMinimize[{OF, cons}, {b, c}, AccuracyGoal -> 4,
PrecisionGoal -> 4, Method -> OptionValue[Method]]
];
```

Εφαρμόστηκε η $FitDistributionTail$ που είναι συνάρτηση των δεδομένων $data$, του πλήθους των σημείων της ουράς $pntsNo$, του είδους της κατανομής (εδώ πρόκειται για την Pareto) και μιας σειράς κανόνων. Τα δεδομένα της ουράς $tailData$ λαμβάνονται ως τα τελευταία $pntsNo$ δεδομένα (όσο είναι δηλαδή το πλήθος των ετών των μετρήσεων του εκάστοτε σταθμού). Ως n ορίζεται το μήκος των δεδομένων ουράς. Παράλληλα τίθενται κάποιοι περιορισμοί στις

παραμέτρους ώστε να έχουν νόημα (*cons*). Το σφάλμα ER είναι συνάρτηση των δεδομένων x και της εμπειρικής κατανομής υπέρβασης $F1$, και ο τελικός τύπος σφάλματος OF δίνεται βάσει της τέταρτης νόρμας (Σχέση 4.17). Για την εύρεση των καταλληλότερων παραμέτρων, ελαχιστοποιείται το σφάλμα λαμβάνοντας υπόψη τους περιορισμούς των παραμέτρων.